Journal of Mathematics and Physics

(Founded by C. L. E. Moore)

6

Volume XI

~~1931~~-1932

Editors

VANNEVAR BUSH
FREDERICK G. KEYES
DIRK J. STRUIK
MANUEL S. VALLARTA
NORBERT WIENER
FREDERICK S. WOODS

Managing Editor
PHILIP FRANKLIN

Published by the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY PRESS

QA1
J87
v.11

OHIO STATE
UNIVERSITY

CONTENTS

PAGE

C. L. E. Moore 1
By D. J. Struik

Dual Pfaffians 12
By C. L. E. Moore and P. Franklin

On the Time Admittances of Transmission Networks . . 27
By John L. Barnes and Clarence T. Prendergast

The Inverse Matrix for De La Vallée-Poussin Summation . 73
By George Rutledge

Synthesis of Electric Networks by Means of the Fourier Transforms of Laguerre's Functions 83
By Y. W. Lee

Uplift Pressure on Dams 114
By W. Weaver

The Rank of a Matrix, the Value of a Determinant, and the Solution of a System of Linear Equations . . . 146
By L. H. Rice

A Recent Contribution to the Design of Electric Filter Networks 150
By E. A. Guillemin

Corrections to the Paper " The Problem of Plateau for Two Contours " by J. Douglas 212

The Field Due to Two Equally Charged Parallel Conducting Cylinders 213
By H. Poritsky

Ritz's Electrodynamic Theory 218
By W. Hovgaard

Systems of Three Linear Partial Differential Equations of the Third Order 255
By G. P. Wadsworth

514468

C. L. E. Moore

C. L. E. MOORE

By D. J. Struik

Clarence Lemuel Elisha Moore was born May 12, 1876, son of George Taylor and Lydia Ann (Bradshaw) Moore, in the town of Bainbridge, Ohio, where his father was a grain dealer. The family settled in Ohio in early pioneer days. Self initiative led him to study first at Ohio State University, then at Cornell. At Ohio State he got his B.Sc. in 1901, at Cornell his A.M. in 1902, his Ph.D. in 1904. A study trip to Europe brought him to Göttingen, Bonn and Turin, where he underwent the lasting influence of Study and of Segre. He joined the teaching staff at Massachusetts Institute of Technology in 1904, and was, since 1909, a member of the faculty, successively as assistant professor, associate professor and full professor. He died December 5, 1931, after a surgical operation. He is survived by his wife, Belle Pease Fuller, whom he married June 11, 1913, and a daughter, Hazel Ann.

He was a Fellow of the American Academy of Arts and Sciences, a member of the American Mathematical Society, Mathematical Association of America, and the Circolo Matematico di Palermo.

In Moore the Massachusetts Institute of Technology loses an excellent educator, the United States an important geometer, and students and colleagues a faithful friend. On the administrative side he was for eleven years in charge of the Institute Course IX (General Science), which includes among its options an undergraduate course in mathematics. He also conducted classes in Theoretical Aeronautics and Advanced Wing Theory.

However, his most important contribution to the educational work of the Massachusetts Institute of Technology came through his interest in mathematical research. Mathematical activity at the Massachusetts Institute of Technology has increased greatly

1

during the last twenty years, a development which is due in a large degree to Moore's own researches and his influence on others. His conception of research was always that of a living, invigorating thing. He held that a mathematician should form the habit of constant production and publication. In this he was in agreement with Faraday, whose maxim was "write, finish, publish." He enforced this rule upon himself though he was for years seriously handicapped by bad eyesight. He also believed, and practised, the precept that one should work at the problems in which he is interested, and which appear valuable, irrespective of whether these problems happen to be popular among mathematicians. His wisdom in this respect is shown by the fact that he was among the earlier workers in fields whose fundamental importance is now universally recognized. Third, and perhaps most significant, he knew that the habit of mind built up within the man himself by the doing of research work is of profound significance in reconstructing the whole mentality, and that a man who had never done any research—regardless of its permanent scientific value—could not be the best type of teacher in a scientific institution. Thus Moore was much given to measuring men and women by their earnestness in research work, by no means to exclusion of other qualities, but as an integral part of their mental make-up. He would sometimes point out how this or that man had been "made over" into a much bigger man by patiently working away on an investigation.

These conceptions he instilled into his associates and pupils, line upon line, with the simplicity, directness, and friendliness, which were his dominating characteristics. People liked to talk things over with Moore, and many an educational or scientific opinion has matured in discussion with him. He was a man, not ignorant of his accomplishments, nor indifferent to them, yet modest enough to be free from vanity. At all times ready with appreciation and praise for the work of others, he was not without criticism of faults and errors, but always in a kind and helpful way. In his dealings with students and younger men

he showed a sympathetic understanding of the character and aspirations of youth rarely found in older men, due in great degree to the simplicity and candor of his own character. In controversies he was willing to put up a fight, but his enmity never became embittered. And if willingness to recognize one's own errors is an indication of greatness of character, Moore certainly possessed this greatness.

Not only as educator was Moore connected with the Massachusetts Institute of Technology, but also as editor and prime mover of this Journal. There he was again induced by his willingness to support younger men, whom he liked to help in a prompt publication of their researches. The products of his own later research are all published in this Journal.

Moore's mathematical work shows a remarkable unity, combined with intuition for living problems and elegance of geometrical conception. Starting from problems on algebraic geometry in ordinary space, following Klein and Segre, he soon became interested in spaces of more dimensions, then in the differential geometry of such spaces, and at last in the differential geometry of Riemannian manifolds. With Ricci and Study, for whom he had also a great admiration, he shared the Leibnizian conviction of the importance of the formal side of the computation, and to this we owe work (done partly in collaboration with H. B. Phillips) on vector analysis and related subjects.

Some of his major contributions to geometry are the investigation of circle complexes, of minimal planes, his theory (together with Phillips) of linear distance and angle in projective geometry, his theory (together with E. B. Wilson) of two dimensional surfaces in hyperspace, his study of remarkable surfaces in hyperspace (rotation surfaces, minimal surfaces, developable surfaces) and of interesting curves in manifolds lying in hyperspace. A series of investigations (together with Philip Franklin) on the geometric properties of Pfaffians occupied his mind at the time when he was stricken.

Though deeply interested in those modern investigations which

tend to broaden the concept of differential geometry by the study of different displacements, Moore himself devoted the time of his research to the study of the many unknown properties of euclidean and Riemannian hyperspace. The deep scientific wisdom of this decision becomes only clear to him who realizes how little we know of the detailed aspect of such manifolds, how surprising the results may be, and how important a thorough investigation is not only for differential geometry, but also for topology, calculus of variations and even physics. In this field Moore stands as sturdy a pioneer as his ancestors were in the woods of the Middle West.

(*A*) *Algebraic Geometry.* The center of Moore's interest in this field were systems of spheres and circles. Klein, in his "Höhere Geometrie" had inspired many geometers to research in this direction, and had pointed to the theory of line systems as a way of starting. Spheres are related to lines in three-space, circles to lines in four-space. In his dissertation (No. 1)† Moore took up the task of interpreting the Klein-Weiler classification of quadratic linear complexes in the language of pentaspherical coordinates. This yielded the classification of quadratic spherical complexes. The singular surface for the general case, analogous to Kummer's surface in line geometry, becomes here a surface of order 24 and class 24 with six double sphero-quartics, which had been studied by P. F. Smith, V. Snyder, and others. The investigation consists, then, in the characterization of all possible distributions of elementary divisors of two quadratic forms in six variables. Moore defines here what he calls the Laguerre-surfaces, which are the envelope of planes belonging to a quadratic spherical complex.

This led Moore to the study of systems of circles. Two projective ways of approach are open. One is to relate the ∞^6 circles in ordinary three-space to the ∞^6 straight lines in non-euclidean four-space. The second is to relate them to a six-dimensional manifold of order 5 M_6^5 in euclidean space of nine

† These numbers refer to the list of publications at the end of this paper.

dimensions. Moore followed both ways (Nos. 2, 3, 4). To get the first representation he followed a suggestion of Laguerre and Darboux and related a plane in space to the two null spheres defined by this plane and a given ordinary sphere. This allowed him not only to study systems of ∞^1, ∞^2, ∞^3, ∞^4 and ∞^5 circles in space, but also to investigate line systems in R_4. There his work runs parallel to contemporary studies of Italian geometers of Segre's school, to whom also Moore often refers (see *Enc. Math. Wiss.*, III C7, Segre, p. 965 and 954). Moore's work is strongly descriptive in character; Coolidge (*Transactions A. M. S.*, 15, 1914) has taken up the subject again from an analytic point of view.

The essential difference between the representation in four-dimensional line space and nine-dimensional point space is that the first gives invariance under the G_{10} of Möbius, the second under the G_{15} of Lie. Moore, however, passed from one to the other for another reason. The representation of circles on lines is good for differential properties of the first order as tangency, but when it comes to questions of osculation point geometry has to be preferred. At this point algebraic geometry in R_3 led him to differential geometry in R_9.

Another series of investigations, partly in collaboration with H. B. Phillips, centered around the notion of linear distance and angle. Ordinary distance is the square root of a homogeneous quadratic form. It is, however, possible to introduce into projective geometry a notion of distance and angle such that points equally distant from a fixed point lie on a line and lines making a given angle with a fixed line pass through a point (Nos. 7 and 8). Necessary for this is a fixed point F and a fixed line f in the plane instead of the fixed curve of second degree of ordinary non-euclidean geometry. It is also possible to extend this notion to n-space.

There are two cases. In the first case F and f are coincident, in the second case they are not. Moore studied also the differential geometry in both cases (No. 10) and found that the first

case is that of the minimal plane, the second that of the minimum developables. There he came to results touching those of Study (*Transactions A. M. S.*, 10, 1909) and Beck (*Archiv für Math. u. Phys.*, 20, 1913, see also L. Berwald, *ib.*, 26, 1915, p. 211).‡

Later Moore came back to this subject (Nr. 27) and determined all manifolds of two and three dimensions in which the differential arc length is a linear function of the parameters. As an example of the results we mention the theorem that such a manifold of three dimensions cannot be realized in euclidean four-space, but only in five-space. This work has been generalized (independently) by J. Lense and others (*Wien. Ber.*, 135, 1926, p. 29).

Several papers of Moore deal with rotations in space of more dimensions (Nr. 14, 16, 19). There he studied, also partly in collaboration with H. B. Phillips, the fundamental classification of rotations and the different type of invariant planes. The main merits of these investigations lie not so much in this classification (this had already been done in several other papers of which we may mention, as not mentioned by the authors, those of Kronecker and Frobenius, see Pascal's Repertorium II, p. 129, art. A. Loewy) as in the study of surfaces invariant under these rotations. This gave Moore a second transition into the differential geometry of more dimensions.

(*B*) *Direct Methods.* Before we discuss this differential geometry we have to deal with the work he (together with H. B. Phillips and E. B. Wilson) did in building up the calculus of vectors and more general quantities for other manifolds than the three-dimensional euclidean ones. (Nos. 7, 8, 14, 15, 16, 19, 20, 28, a paper 23 on a hypercomplex number system is related). These authors, following Grassmann and Gibbs, extend the dyadic calculus of Gibbs to n dimensions and to higher dyadics than those of type one–one. Moore used his generalized calculus for the classification of rotations in hyperspace. He also used

‡ H. Berliner came independently to a related geometry (Math. Annalen 79, 1919, p. 13).

the Grassmann calculus for his investigations of plane geometry. All these attempts, however, are subject to the criticism of Klein and Schouten that the central classification principle of the transformation group is missing, so that the complete system is not found, and to that of Study, stating that the completeness of the system is not proved. Moore, in company with many other workers in the same field, might have felt this himself when in his later work on differential geometry he applied the system of Ricci, which seems to realize all the advantages of the direct methods by doing entirely away with them.

(*C*) *Differential Geometry.* Moore's work in this branch of science starts with his study of the M_6^5 in R_9, derived from circle geometry. His investigations always show two sides, the study of curves and manifolds characterized by general properties, as conjugate lines, asymptotic lines, curvature manifolds, or minimal varieties, combined with the study of special manifolds which serve as examples. We mention the following results.

Moore found the asymptotic lines of higher order on an m-dimensional manifold in R_n. These are curves of which the osculating R_s lies in the tangent R_m (Nos. 5, 6). In connection with this he defined conjugate directions for such manifolds and their dual complements in the sense of Segre (*Rend. Lincei* 17, 1908, p. 405).

Moore (in collaboration with E. B. Wilson) developed a general theory of two-dimensional manifolds in euclidean n-space (Nos. 9, 11, 12). This theory was already developed to a certain extent by E. E. Levi (*Ann. Sc. Norm. Pisa*, 10, 1905). Moore and Wilson took up the study again with the consistent use of Ricci's absolute differential calculus. The elementary curvature properties of such surfaces are determined by a curvature cone of the second order lying in a three-dimensional space perpendicular to the tangent plane. This is the indicatrix. It allows an easy study of the general invariant quantities, as e.g., the mean curvature vector. When this vector vanishes, the surface is a minimal surface. Other surfaces can be characterized

in a similar way as developable surfaces, surfaces with axial points, planar points, ruled surfaces and special directions as asymptotic directions and principal directions. The paper is still the last word on the subject (see *Enc. Math. Wiss.*, III D, 11, art. L. Berwald).

Connected with his investigations on rotations in hyperspaces is a series of papers on two-dimensional rotation surfaces and translation surfaces in such spaces (Nos. 17, 18, 21), and their connection with developable and minimal varieties. Among the results we mention the theorem that minimal surfaces are translation surfaces, and a necessary and sufficient condition that a surface in hyperspace be a minimum surface. This condition is that the minimal lines on it are characteristics. (Similar results were found at the same time by Bompiani, *Comptes Rendus* 169, p. 840, 1919). Rotation surfaces of constant curvature in R_4 received special attention.

In a study on developable surfaces (Nr. 26) Moore gave his attention to certain types of such surfaces which are not ruled, especially those which project on $n - 2$ mutually perpendicular three-spaces into ordinary developables.

Moore also studied under what conditions normal planes to a surface in R_4 will envelop a second surface, and the related question of parallel surfaces (No. 24).

Several papers deal with the application of covariant differentiation and Levi-Civita's parallelism to more dimensional manifolds not only euclidean, but also Riemannian (Nos. 25, 29, 30, 31, 32, 33). There we find studies of parallel geodesics, on Codazzi equations, on congruences of curves. The methods of Segre, who liked to investigate manifolds in higher dimensions with the aid of partial differential equations, were also favorite with Moore. He showed (No. 29) how these methods can be applied to Riemannian geometry or geometry of more general displacement. He also, in collaboration with Ph. Franklin (No. 35) began to study such Segre systems analytically, as a means of acquiring more detailed information on manifolds of more than

two dimensions, in a direction where so far only Sisam (*Amer. Journ.*, 33, p. 97, 1911) has been a pioneer.

During the last years of his life Moore took up, in collaboration with Philip Franklin, the geometrical aspects of Pfaff equations, which yield a generalization of the theory of manifolds (Nos. 34, 36, 37, 38, 39).

It is a pity that Moore never attempted to put his results on more dimensional differential geometry together. Just through his persistent efforts to pass from the general to the concrete a book from his hand would have been the necessary supplement to the textbooks on more dimensional differential geometry that have appeared so far, and that all emphasize the general theorems. The leading thought seems sometimes less the development of the geometrical entity than that of the formal calculus. Now we can only hope that future authors will seriously study the lessons that Moore had to teach.

PUBLICATIONS OF C. L. E. MOORE

1. Classification of Surfaces of Singularities of the Quadratic Spherical Complex. Diss. Cornell, 1904. *Amer. Jour.*, 27, pp. 248–279, 1905.
2. Circles Orthogonal to a Given Sphere. *Annals of Math.*, 2d ser., V, pp. 57–72, 1907.
3. Infinitesimal Properties of Lines in S_4 with Application to Circles in S_3. *Proc. Amer. Acad.*, 46, pp. 343–362, 1911.
4. Some Properties of Lines in Space of Four Dimensions and their Interpretation in Circle Geometry in Space of Three Dimensions. *Amer. Journal*, 33, pp. 129–152, 1911.
5. Conjugate Directions on a Hypersurface in a Space of Four Dimensions and Some Allied Curves. *Ann. of Math.*, 2d ser., 13, pp. 89–102, 1911.
6. Surfaces in Hyperspace which Have a Tangent Line with Three-Point Contact Passing through Each Point. *Bull. Amer. Math. Soc.*, 18, pp. 284–290, 1912.
7. An Algebra of Plane Projective Geometry. Co-author H. B. Phillips. *Proc. Amer. Acad.*, 47, pp. 735–790, 1912.
8. A Theory of Linear Distance and Angle. Co-author H. B. Phillips. *Proc. Amer. Acad.*, 48, pp. 43–80, 1912.
9. Note on Normal Sections of a Surface in a Space of n-Dimensions. *Ann. Math.*, 16, pp. 89–96, 1915.
10. Geometry whose Element of Arc is a Linear Differential Form, with Applications to the Study of Minimum Developables. *Proc. Amer. Acad.*, 50, pp. 197–222, 1915.

11. A General Theory of Surfaces. *Proc. Nat. Acad.*, 2, pp. 273–278, 1916.
12. Differential Geometry of Two-Dimensional Surfaces in Hyperspace. Co-author E. B. Wilson. *Proc. Amer. Acad.*, 52, pp. 267–368, 1916.
13. A Substitute for Dupin's Indicatrix. *Amer. Math. Monthly*, 24, pp. 456–462, 1917.
14. Motions in Hyperspace. *Ann. of Math.*, 19, pp. 176–184, 1918.
15. The Dyadics which Occur in a Point Space of Three Dimensions. Co-author H. B. Phillips. *Proc. Amer. Acad.*, 53, pp. 387–438, 1918.
16. Rotations in Hyperspace. *Proc. Amer. Acad.*, 53, pp. 649–694.
17. Translation Surfaces in Hyperspace. *Bull. Amer. Math. Soc.*, 25, pp. 75–85, 1918.
18. Surfaces of Rotation in a Space of Four Dimensions. *Ann. Math.*, 21, pp. 81–93, 1919.
19. Rotations in Spaces of Even Dimensions. Co-author H. B. Phillips. *Proc. Amer. Acad.*, 55, pp. 157–188, 1920.
20. Note on Geometrical Products. Co-author H. B. Phillips. *Proc. Nat. Acad.*, 6, pp. 155–158, 1920.
21. Rotation Surfaces of Constant Curvature in a Space of Four Dimensions. *Bull. Amer. Math. Soc.*, 26, pp. 454–460, 1920.
22. Note on Minimal Varieties in Hyperspace. *Bull. Amer. Math. Soc.*, 27, pp. 211–216, 1921.
23. Hyperquaternions. *Jour. Math. and Physics*, 1, pp. 63–77, 1922.
24. Note on the Normal Planes to a Surface in a Space of Four Dimensions. *Jour. Math. and Physics*, 1, pp. 147–156, 1922.
25. Note on the Vanishing of the Determinant of the Second Fundamental Form of a Hypersurface. *Jour. Math. and Physics*, 2, pp. 125–130, 1923.
26. Note on Developable Surfaces in Hyperspace. *Jour. Math. and Physics*, 3, pp. 1–6, 1924.
27. Minimal Varieties of Two and Three Dimensions whose Element of Arc is a Perfect Square. *Jour. Math. and Physics*, pp. 167–178, 1925.
28. Grassmannian Geometry in Riemannian Space. *Jour. Math. and Physics*, 5, pp. 191–200, 1926.
29. Note on Surfaces in a non-Riemannian Space. *Jour. Math. and Physics*, 6, pp. 39–49, 1926.
30. Parallel Geodesics. *Jour. Math. and Physics*, 7, pp. 81–85, 1928.
31. Note on Congruences in Riemannian Space. *Jour. Math. and Physics*, 7, pp. 185–192, 1928.
32. Principal Directions in a Subspace V_m Immersed in a Riemannian Space V_n. *Jour. Math. and Physics*, 8, pp. 65–69, 1929.
33. Hypersurfaces for which the Codazzi Equations Take the Simplest Form. *Jour. Math. and Physics*, 8, pp. 141–147, 1929.
34. Geometry of a Complex of Curves in a Riemannian Space of n Dimensions. *Jour. Math. and Physics*, 8, pp. 247–265, 1929.
35. Systems of Linear Partial Differential Equations. Co-author P. Franklin. *Jour. Math. and Physics*, 9, pp. 22–46, 1930.

36. Differential Projective Geometry of a System of Pfaffian Differential Equations. *Jour. Math. and Physics*, 9, pp. 320–332, 1930.
37. The Geometry of Algebraic Pfaffians. Co-author P. Franklin. *Jour. Math. and Physics*, 10, pp. 19–49, 1931.
38. Pfaffians in Parametric Forms. Co-author P. Franklin. *Jour. Math. and Physics*, 10, pp. 95–105, 1931.
39. Geodesics of Pfaffians. Co-author P. Franklin. *Jour. Math. and Physics*, 10, pp. 157–190, 1931.
40. Dual Pfaffians. Co-author P. Franklin. *Jour. Math. and Physics*, 11, pp. 12–26, 1932.

DUAL PFAFFIANS[1]

By C. L. E. Moore and P. Franklin

1. *Introduction.* Our previous studies have dealt with certain geometric properties of a Pfaffian in Euclidean 3-space, and to a limited extent with their generalizations for systems of m Pfaffians in Riemannian n-space.[2] The present paper is devoted to a discussion of duality relations in Riemannian n-space, and their application to problems involving Pfaffian configurations. These duality relations exist between the elements of the linear vector n-space associated with each point of a Riemannian n-space. In defining them, we make use of a new type of co-ordinate for m dimensional vector sub-spaces of linear vector n-spaces which has some analogies to Plückerian line co-ordinates. After defining dual sub-spaces, and the system of equations dual to a set of Pfaffian relations, we are in a position to formulate the dual of a minimization problem, such as that of finding the geodesic m-volumes in Pfaffian $(m + k)$ dimensional sub-spaces.

2. *Plückerian Co-ordinates and Dual Sub-spaces.* At each point of a Riemannian n-space the differential vectors form a linear vector n-space. These vectors lie in an n-dimensional affine point space, which gives rise to an $(n - 1)$ dimensional projective point space at infinity. The unit vectors may be characterized by the points of this latter $(n - 1)$ space. An m-dimensional vector sub-space is obtained by taking all the vectors dependent on m linearly independent ones. As the same sub-space results from any m linearly independent vectors con-

[1] Presented to the American Mathemaicl Society, September, 1931. Owing to the subsequent death of Professor Moore, I must take full responsibility for any shortcomings in the final editing. P. F.

[2] Cf. (I) The Geometry of Algebraic Pfaffians, vol. X, p. 19; (II) Pfaffians in Parametric Form, vol. X, p. 95; and (III) Geodesics of Pfaffians, vol. X, p. 157; this Journal, together with the references there given to other papers in this field.

tained in it, there is a wide choice in these. Let the parameters associated with the linearly independent vectors be $u^1, u^2, \cdots u^m$, so that the components of the jth vector are

$$(1) \qquad dx^1/du^j,\ dx^2/du^j,\ \cdots\ dx^n/du^j, \quad (j = 1, 2, \cdots m).$$

These quantities make up a matrix of m rows and n columns, and our new type of co-ordinate for the sub-space is obtained by selecting the m-rowed square minors of this matrix:

$$(2) \qquad a^{i_1 i_2 \cdots i_m} = |dx^{i_k}/du^j| \quad (k, j = 1, 2, \cdots m).$$

We use each combination only once, so that the co-ordinates are

$$ {}_nC_m = \frac{n!}{m!(n-m)!},$$

in number.

There are usually a number of identical relations satisfied by these co-ordinates. Thus, if $2m \leq n$, for any $2m$ distinct integers between 1 and n, $i_1, i_2, \cdots i_{2m}$, the determinant

$$|dx^{i_k}/du^s| = 0, \quad \begin{matrix}(k = 1, 2, \cdots 2m),\\ (s = 1, 2, \cdots m, 1, 2, \cdots m),\end{matrix}$$

since the last m rows are identical with the first m. An expansion of this by Laplace's development [3] leads to the identity:

$$(3) \qquad \sum \pm a^{p_1 p_2 \cdots p_m} a^{q_1 q_2 \cdots q_m} = 0,$$

where $p_1 p_2 \cdots p_m q_1 q_2 \cdots q_m$ is a separation of $i_1 i_2 \cdots i_{2m}$, each combination being counted only once, and the sign being plus or minus according as the permutation written is even or odd.

When $2m > n$, a similar set of identities may be obtained by using the relations of these quantities to the $(n - m)$ rowed minors of the adjoint determinant,[4] since then $2(n - m) < n$. We take as the fundamental determinant of the adjoint relations one whose first m rows are the quantities given in (1) above, and whose remaining rows are taken to make the determinant unity, which is possible since the original vectors are linearly inde-

[3] Bôcher, *Introduction to Higher Algebra*, New York, 1907, p. 24.
[4] Bôcher, l. c. 3, p. 30.

pendent. The identities may also be obtained more directly by writing down a determinant of the $2m$th order with $(2m - n)$ columns repetitions of earlier columns, and expanding as before. Care must be taken that each two identical columns appear in even, or both in odd places when m is even, while they must appear in places of unlike parity when m is odd. For, if the identical columns are the ith and jth, when $i + j + m$ is even, each term of the identity will appear twice, with the same sign, but if the sum is odd, the terms will appear twice with opposite sign, and so will all cancel out. The identity is similar in form to (3), except that the terms and signs are obtained from the separations of $i_{2m+1}, i_{2m+2}, \cdots i_{2n}$ into $p_{m+1}'p_{m+2}', \cdots p_n'q_{m+1}'q_{m+2}', \cdots q_n'$, where the primed sets are the complements of the corresponding unprimed sets appearing in the identity.

In general, the identities just mentioned will not all be independent. For example, for $m = 1$, they become trivial. The precise number of independent identities may be calculated by noting that there are ${}^nC_m$ co-ordinates. Their absolute magnitudes are immaterial so that the ratios, which alone enter into the homogeneous identities, are one less than this in number. Finally an m dimensional sub-space in an n dimensional vector space is specified by the same number of parameters as are required to specify an $(m - 1)$ dimensional point space in $(n - 1)$ dimensions, namely $m(n - m)$. Thus there must be

$$I = {}^nC_m - 1 - m(n - m)$$

independent identities. For example, when $n = 4$, $m = 2$ there is just one relation. This is the case of planar vector elements in the vector 4 space, which lead to lines in 3 space for the infinite region. Our co-ordinates are essentially the ordinary Plücker co-ordinates of these lines. For this reason we have called them Plückerian even in the general case.

If we determine a given m-dimensional sub-space by two different sets of independent vectors, their co-ordinates will give $a^{i_1 i_2 \cdots i_m}$ having the same ratios. For the transition from one

set of vectors to the other may be accomplished by a series of simpler transformations, each of the latter being such that it leaves all but one of the vectors unchanged, and replaces this one by a linear combination of itself (with a non-zero coefficient) and the other vectors. But such transformations will evidently at most change the $a^{i_1 i_2 \cdots i_m}$ by a common factor.

To show, conversely, that when two sub-spaces S, S' have the same ratio of their $a^{i_1 i_2 \cdots i_m}$ they are the same, we note that our Plückerian co-ordinates have tensor character. But, if we change to normal co-ordinates such that the new first m co-ordinate vectors are an orthogonal entuple in the S sub-space, that is, each is a linear combination of the first set of vectors used to determine S, we shall have all the Plückerian co-ordinates zero except

$$a^{1,2\cdots m} \neq 0.$$

Since S' had the same ratio of co-ordinates in the first system, from their tensor character it will have the same ratio of co-ordinates for the new system. Hence no vector in S' has a component with index greater than m different from zero, so that S' lies in S, and so must coincide with it.

Our operations on determinants will be simplified by the use of a notation originally due to Eddington.[5] We define

$$\epsilon_{i_1 i_2 \cdots i_n} = \epsilon^{i_1 i_2 \cdots i_n} = 0, \text{ if two } i\text{'s are equal,} \tag{4}$$
$$= 1 \text{ for an even permutation,}$$
$$= -1 \text{ for an odd permutation.}$$

Thus, with the summation convention, this makes any determinant

$$|a_{jk}| = a_{1, i_1} a_{2, i_2} \cdots a_{n, i_n} \epsilon^{i_1 i_2 \cdots i_n}.$$

We observe that

$$\epsilon_{i_1 i_2 \cdots i_n} \sqrt{g} a_1{}^{i_1} a_2{}^{i_2} \cdots a_n{}^{i_n} = \sqrt{g} |a_j{}^{i_k}|,$$
$$(\epsilon^{i_1 i_2 \cdots i_n}/\sqrt{g}) g_{i_1 j_1} g_{i_2 j_2} \cdots g_{i_n j_n} a_1{}^{j_1} a_2{}^{j_2} \cdots a_n{}^{j_n} = \sqrt{g} |a_j{}^{i_k}|. \tag{5}$$

[5] Eddington, *The Mathematical Theory of Relativity*, Cambridge, 1923, p. 107. See also Levi-Civita, *Lezioni di calcolo differenziale assoluto*. Rome 1925, p. 180 and Eisenhart, *Riemannian Geometry*, Princeton, 1926, p. 101.

Since the right member of these equations is a well known expression for the volume of the parallelopiped on the n vectors $a_i{}^a$, and hence a scalar, it follows that the quantities

$$\epsilon_{i_1 i_2 \cdots i_n}\sqrt{g} \quad \text{and} \quad \epsilon^{i_1 i_2 \cdots i_n}/\sqrt{g} \tag{6}$$

are, respectively, a covariant and a contravariant tensor.

We also define

$$\delta^{i_1 i_2 \cdots i_m}_{j_1 j_2 \cdots j_m} = 0, \text{ unless the } i\text{'s and the } j\text{'s are permutations of the same combination of } m \text{ distinct values,}$$
$$= 1, \text{ for an even permutation,}$$
$$= -1 \text{ for an odd permutation.} \tag{7}$$

That this is a mixed tensor, as indicated by the subscripts, follows from the fact that

$$\delta^{i_1 i_2 \cdots i_m}_{j_1 j_2 \cdots j_m} = |\delta^{i_p}_{j_q}|, \quad (p, q = 1, 2, \cdots m), \tag{8}$$

a determinant built up from the Kronecker delta, which is known to have tensor character.

The identity (3) previously found for the Plückerian coordinates may now be written more precisely as:

$$\delta^{p_1 \cdots p_m q_1 \cdots q_m}_{i_1 \cdots i_{2m}} a^{p_1 \cdots p_m} a^{q_1 \cdots q_m} = 0, \quad n \geq 2m,$$
$$\delta^{p' m+1 \cdots p' n q' m+1 \cdots q' n}_{i_{2m+1} \cdots i_{2n}} a^{q_1 \cdots p_m} a^{q_1 \cdots p_m} = 0, \quad n < 2m. \tag{9}$$

In the second equation, as before, p' and q' denote elements of the sets complementary to those of the p and q.

In a projective space of n dimensions, the ordinary duality relations are those between point spaces of m and $n - m - 1$ dimensions. Each of these spaces depends on $(m + 1)(n - m)$ parameters. Since the n-dimensional vector spaces with which we are concerned here may be associated with $(n - 1)$ dimensional point spaces at infinity, the vector duality relations will correspond to those between $(m - 1)$ and $(n - 1) - (m - 1) - 1$ dimensional point spaces, and hence be between m and $(n - m)$ dimensional vector spaces. These will each depend on $m(n - m)$ parameters.

We now define the Plückerian co-ordinates of the $(n-1)$ dimensional vector space dual to a vector by the equations:

$$A^{i_1\cdots i_n} = \epsilon^{i_1\cdots i_n} g_{i_1 p} a^p / \sqrt{g}. \tag{10}$$

More generally, for the $(n-m)$ dimensional dual of an m-dimensional sub-space, we put:

$$A^{i_{m+1}\cdots i_n}_{\cdot} = \epsilon^{i_1\cdots i_n} g_{i_1 p_1} \cdots g_{i_m p_m} a^{p_1\cdots p_m} / (\sqrt{g} m!). \tag{11}$$

It follows from this definition that the dual of $A^{i_{m+1}\cdots i_n}$ is given by:

$$b^{i_1\cdots i_m} = \epsilon^{i_1\cdots i_n} g_{i_{m+1} i_{m+1}} \cdots g_{i_n i_n} A^{i_{m+1}\cdots i_n} / (\sqrt{g}(n-m)!)$$
$$= \frac{\epsilon^{i_1\cdots i_n} \epsilon^{i_1\cdots i_n}}{gm!(n-m)!} g_{i_1 p_1} \cdots g_{i_m p_m} g_{i_{m+1} i_{m+1}} \cdots g_{i_n i_n} a^{p_1\cdots p_m}.$$

In this summation, the coefficient for any set of $p_1 \cdots p_m$ which are not a permutation of the $j_1 \cdots j_m$ will be zero. For, in this case, since the $j_{m+1} \cdots j_n$ are not complementary to the $p_1 \cdots p_m$, they must have some common element, say $j_{m+s} = p_t$. This makes the product

$$g_{i_t p_t} g_{i_{m+s} i_{m+s}}$$

preserve its value when i_t and i_{m+s} are interchanged, but since the factor $\epsilon^{i_1\cdots i_n}$ changes its sign with this interchange, the two terms cancel. Similarly all the terms of the coefficient of $a^{p_1\cdots p_m}$ cancel in pairs. When the set $p_1 \cdots p_m$ are a permutation of the $j_1 \cdots j_m$, we get the same value for each of the $m!$ permutations, since an interchange of two p_i changes the sign of $a^{p_1\cdots p_m}$, which is compensated by the factor $\epsilon^{i_1\cdots i_n}$, since the i's and p's must be interchanged together to give the same factor in the g_{ip}. To make the ϵ factors different from zero, the $j_{m+1} \cdots j_n$ must be in the set complementary to the $p_1 \cdots p_m$ and again each of the $(n-m)!$ permutations gives the same result, since to preserve the g_{ji}, an interchange of two j's must be accompanied by an interchange of two i's, and both ϵ's change sign. Finally, we note that in particular for a single choice of $j_{m+1} \cdots j_n$, and for $p_1 \cdots p_m = j_1 \cdots j_m$, the summation for the i's in the coefficient

gives:

$$|g_{ij}| = g,$$

so that the result of the summation is

$$b^{i_1 \cdots i_m} = a^{i_1 \cdots i_m}.$$

We have just demonstrated that *the duality relation, as defined by* (11) *is reciprocal.* If we were only concerned with the ratios of the co-ordinates, the factor $m!$ could have been omitted. It was included so that, if we normalize the co-ordinates by fixing the scale, the duality relation will still be a reciprocal one.

We will next show that *our dual spaces are perpendicular*, in the sense that every vector in a sub-space is perpendicular to every vector in the dual sub-space. For this purpose, we introduce a set of normal co-ordinates such that the first m co-ordinate vectors span the sub-space $a^{i_1 \cdots i_m}$. Then we have, for the proper scale factor:

$$a^{i_1 \cdots i_m} = \delta^{i_1 \cdots i_m}_{1 \cdots m}.$$

Since our definition of duality has tensor character, in these normal co-ordinates the dual space will have as its co-ordinates:

$$\begin{aligned} A^{i_{m+1} \cdots i_n} &= \epsilon^{i_1 \cdots i_n} \delta_{i_1 p_1} \cdots \delta_{i_m p_m} a^{p_1 \cdots p_m}/m! \\ &= \delta^{i_{m+1} \cdots i_n}_{m+1 \cdots n}. \end{aligned}$$

This shows that every vector in the first space is a linear combination of the first m co-ordinate vectors, while every one in the dual space is a linear combination of the last $n - m$ co-ordinate vectors, which proves the relation of perpendicularity.

We may also conclude that if we normalize the co-ordinates to such a scale that the parallelopiped on the vectors defining the sub-space is of unit volume, the same will be true of the dual space. Finally, we note that the equation just written gives a shorter proof of the reciprocal character of our duality relations.

Our definition of dual sub-spaces, (10), has been formulated in such a way as to give the contravariant components of the dual sub-space in terms of the contravariant components for the

original sub-space. An equivalent formulation for covariant components is:

$$A_{i_{m+1}\cdots i_n} = \epsilon_{i_1\cdots i_n} g^{i_1 p_1}\cdots g^{i_m p_m} a_{p_1\cdots p_m} \sqrt{g}/m!, \tag{12}$$

while for changing from covariant components in one sub-space to contravariant components in the dual sub-space, we have the simpler forms:

$$A^{i_{m+1}\cdots i_n} = \epsilon^{i_1\cdots i_n} a_{i_1\cdots i_m}/(\sqrt{g}m!), \tag{13}$$

and

$$A_{i_{m+1}\cdots i_n} = \epsilon_{i_1\cdots i_n} a^{i_1\cdots i_m}\sqrt{g}/m!. \tag{14}$$

It should be observed that our definitions apply in the first instances to sub-spaces. However, if we start with any tensor which is skew-symmetric in all its indices, we may use the definitions to find a dual skew-symmetric tensor. The relation will be reciprocal, since in our first proof of the reciprocal character of duality, the only property of the Plückerian co-ordinates used was their skew-symmetrical character. If we attempted to apply the definitions to tensors with no special symmetry properties, while new tensors would be derived, the relation so obtained would not be a reciprocal one. In the sequel we shall only apply our definitions to skew-symmetric tensors.

3. *Application of Duality to Scalar Relations and Arc Length.* If we have two sub-spaces of the same order, or more generally two skew-symmetric tensors, we may form a scalar from them:

$$m!k = a_{i_1\cdots i_m} b^{i_1\cdots i_m} = a^{i_1\cdots i_m} b_{i_1\cdots i_m}. \tag{15}$$

The corresponding scalar for the dual tensors is:

$$\begin{aligned}(n-m)!K &= A_{i_{m+1}\cdots i_n} B^{i_{m+1}\cdots i_n} \\ &= \frac{\epsilon_{p_1\cdots p_m i_{m+1}\cdots i_n}\epsilon^{q_1\cdots q_m i_{m+1}\cdots i_n}}{(m!)^2} a^{p_1\cdots p_m} b_{q_1\cdots q_m} \\ &= \frac{(n-m)!}{(m!)^2}\delta^{q_1\cdots q_m}_{p_1\cdots p_m} a^{p_1\cdots p_m} b_{q_1\cdots q_m} \\ &= (n-m)!k.\end{aligned} \tag{16}$$

Thus the scalar k is equal to the corresponding scalar K formed for the dual sub-space. We may deduce other expressions for the scalar k from (15). We have:

$$\begin{aligned} k &= a_{i_1\cdots i_m} b^{i_1\cdots i_m}/m! \\ &= \delta^{p_1\cdots p_m}_{i_1\cdots i_m} a_{p_1\cdots p_m} b^{i_1\cdots i_m}/(m!)^2 \\ &= \delta^{p_1\cdots p_m}_{i_1\cdots i_m} g_{i_1 p_1}\cdots g_{i_m p_m} a^{i_1\cdots i_m} b^{i_1\cdots i_m}/(m!)^2 \\ &= \delta^{p_1\cdots p_m}_{i_1\cdots i_m} g^{i_1 q_1}\cdots g^{i_m q_m} a_{p_1\cdots p_m} b_{q_1\cdots q_m}/(m!)^2. \end{aligned}$$

These expressions may be abbreviated by introducing a notation for the multiples of the minors of the determinant of the g which occur. We put:

$$(17) \qquad g_{i_1\cdots i_m,\, j_1\cdots j_m} = \delta^{p_1\cdots p_m}_{i_1\cdots i_m} g_{i_1 p_1}\cdots g_{i_m p_m}/(m!)^2.$$

This leads to:

$$(18) \qquad g^{p_1\cdots p_m}_{i_1\cdots i_m} = \delta^{p_1\cdots p_m}_{i_1\cdots i_m}/(m!)^2,$$

and

$$(18) \qquad g^{i_1\cdots i_m,\, j_1\cdots j_m} = \delta^{i_1\cdots i_m}_{p_1\cdots p_m} g^{j_1 p_1}\cdots g^{j_m p_m}/(m!)^2.$$

The last two expressions for k may then be written:

$$(19) \qquad \begin{aligned} k &= g_{i_1\cdots i_m,\, j_1\cdots j_m} a^{i_1\cdots i_m} b^{j_1\cdots j_m} \\ &= g^{i_1\cdots i_m,\, j_1\cdots j_m} a_{i_1\cdots i_m} b_{j_1\cdots j_m}. \end{aligned}$$

The corresponding equal scalar for the dual sub-space, K, of (16), is similarly transformed into:

$$(20) \qquad \begin{aligned} K &= g_{i_{m+1}\cdots i_n,\, j_{m+1}\cdots j_n} A^{i_{m+1}\cdots i_n} B^{j_{m+1}\cdots j_n} \\ &= g^{i_{m+1}\cdots i_n,\, j_{m+1}\cdots j_n} A_{i_{m+1}\cdots i_n} B_{j_{m+1}\cdots j_n}. \end{aligned}$$

Since the sub-spaces here considered are uniquely determined by their dual sub-spaces, and vice versa, instead of regarding the duals of the Plückerian co-ordinates of a sub-space as corresponding to a new geometric entity, we could regard them as a new type of co-ordinate for the original sub-space. We shall for the most part adapt the first point of view. The second interpretation might be useful in referring the formulas and properties of

sub-spaces of order m with m greater than $n/2$ to those of order m' with m' less than $n/2$.

The ordinary arc element:

$$(21) \qquad ds^2 = g_{ij}dx^i dx^j = dx^i dx_i,$$

is a scalar similar to the k of (15) or (19) formed for a one dimensional sub-space. For the $(n-1)$ space dual to this vector, we form the element:

$$(22) \qquad ds_{n-1}^2 = dX_{i_2 \cdots i_n} dX^{i_2 \cdots i_n}/(n-1)! = g_{i_2 \cdots i_n, j_2 \cdots j_n} dX^{i_2 \cdots i_n} dX^{j_2 \cdots j_n}.$$

We have seen that this is equal to ds^2.

For an m-dimensional sub-space, we have as the element of volume:

$$(23) \qquad ds_m^2 = dx_{i_1 \cdots i_m} dx^{i_1 \cdots i_m}/m! = g_{i_1 \cdots i_m, j_1 \cdots j_m} dx^{i_1 \cdots i_m} dx^{j_1 \cdots j_m}.$$

The dual of this is:

$$(24) \qquad ds_{n-m}^2 = dX_{i_{m+1} \cdots i_n} dX^{i_{m+1} \cdots i_n}/(n-m)! = g_{i_{m+1} \cdots i_n, j_{m+1} \cdots j_n} dX^{i_{m+1} \cdots i_n} dX^{j_{m+1} \cdots j_n}.$$

For two distinct vectors, we may form the scalar product as the bilinear form corresponding to the quadratic form which give the arc length. Similarly, for two distinct sub-spaces, we may form an m-dimensional scalar product:

$$(25) \qquad \cos(s, t) ds_m dt_m = dx_{i_1 \cdots i_m} dy^{i_1 \cdots i_m}/m! = g_{i_1 \cdots i_m, j_1 \cdots j_m} dx^{i_1 \cdots i_m} dy^{j_1 \cdots j_m}.$$

The form of the left member of this equation is intended to suggest that this scalar product is the product of the volumes spanned by the basic vectors of the two spaces, multiplied by a factor less than unity.[6] The factor less than one, the cosine of the " angle "

[6] Cf. S. Ganguli, *Bulletin of the Calcutta Math. Soc.*, vol. IX, 1918, p. 11, and for the tensor form a paper by one of the writers in the same journal, vol. XIII, 1922, p. 47. We may also refer to S. Ganguli, *Analytical Geometry of Hyper-spaces*, University of Calcutta, 1918, where earlier papers are cited.

between the spaces, is equal to the product of the m extremal values of the cosine of an angle between two vectors, one in each space. From the fact that scalars formed from two sub-spaces are the same as those formed from the dual sub-spaces, we see that on taking the duals the scalar product, as well as the volume elements are unchanged. Consequently, *the angle between two sub-spaces is the same as the angle between their duals.*

For later use, we recall some considerations on geodesic sub-spaces.[7] In n space, a closed $m-1$ dimensional hypersurface bounds an m dimensional hypervolume, which may be computed by taking the m-tuple integral of the expression for the differential of volume already given, (23). It will be convenient to introduce parameters, $u^1, u^2, \cdots u^m$ in the sub-space. We regard the $a^{i_1 \cdots i_m}$ as given in terms of derivatives dx^k/du^i, and write the earlier

$$dx^{i_1 \cdots i_m} = a^{i_1 \cdots i_m} du^1 \cdot\cdot\cdot du^m. \tag{26}$$

Then the m-dimensional volume is given by:

$$I_m = \int_V \sqrt{g_{i_1 \cdots i_m, j_1 \cdots j_m} a^{i_1 \cdots i_m} a^{j_1 \cdots j_m}} du^1 \cdots du^m, \tag{27}$$

where the integral is to be taken over the m-volume bounded by the $(m-1)$-surface. If we denote the integrand by L, we may express the conditions for an extremal in the form:

$$\partial/\partial u^i(\partial L/\partial x^p{}_{,i}) - \partial L/\partial x^p = 0, \quad (p = 1, 2, \cdots n), \quad (i = 1, 2, \cdots m). \tag{28}$$

4. *Pfaffian Sub-spaces.* Before applying the notion of duality to Pfaffian expressions, we must explain the interpretation of sub-spaces satisfying Pfaffian relations in terms of their Plückerian co-ordinates. A single Pfaffian expression set equal to zero, as

$$A_i dx^i = 0, \quad (i = 1, 2, \cdots n) \tag{29}$$

is satisfied by vectors, dx^i, perpendicular to the vector A^i.

[7] Eisenhart, *Riemannian Geometry*, Princeton, 1926, p. 177.

Accordingly it defines an $n-1$ dimensional element at each point of the n-space in which we are working. If several such relations are given,

$$(30) \qquad {}^{p}A_{i}dx^{i} = 0, \quad (p = 1, 2, \cdots m),$$

any vector satisfying all of them simultaneously must be perpendicular to the space spanned by the ${}^{p}A_{i}$, and hence be in the space dual to it. Thus *a system of m linearly independent Pfaffian equations defines an $n-m$ space.* The Plückerian co-ordinates for this are readily found from the coefficients of the m Pfaffians.

We next consider a relation linear in the Plückerian co-ordinates of a two dimensional element, as:

$$(31) \qquad A_{ij}dx^{ij} = 0.$$

We may without loss of generality restrict the coefficients of an expression like this to be skew-symmetric. This follows from the skew-symmetrical character of the dx^{ij}, which makes the form just written equal to a second one with:

$$A_{ij}' = \tfrac{1}{2}(A_{ij} - A_{ji}).$$

The equation (31) states that any two dimensional element satisfying it is perpendicular to the element A^{ij}, and hence lies in the dual $n-2$ space. If a series of such Pfaffians, say m of them, is given and the elements corresponding to the coefficients are all perpendicular to one another, they will span a $2m$ space. The Plückerian co-ordinates of this $2m$ space may be taken as suitable mth order determinants in the ${}^{p}A^{ij}$, the contravariant components of ${}^{p}A_{ij}$, p being an index running from 1 to m. In this case, any two dimensional element satisfying the system of Pfaffians must lie in the dual $n-2m$ space. In general, the spaces of the ${}^{p}A_{ij}$ will have common directions, so that they will span a space of lower order than the $2m$th, and the system will not restrict the dx^{ij} to this extent.

Similar considerations apply to the system of equations:

$$(32) \qquad {}^{p}A_{i_1 \cdots i_m}dx^{i_1 \cdots i_m} = 0. \qquad (m \leq n/2)$$

Each of these equations restricts an m dimensional element to lie in an $n - m$ dimensional sub-space. If there are r such relations, they will restrict the m dimensional element to lie in a space of order M, where

$$n - rm \leq M \leq n - m. \tag{33}$$

The dual of each Pfaffian of the system just discussed will be of the form:

$$^{p}a_{i_{m+1}\cdots i_n} dX^{i_{m+1}\cdots i_n} = 0, \tag{34}$$

and will restrict the dual $n - m$ dimensional element to contain an m dimensional sub-space. The complete system will restrict the dual $n - m$ dimensional element to contain an M dimensional sub-space.

To set up minimum problems in Pfaffian sub-spaces of our n-space, we assume that we have given a fundamental M dimensional Pfaffian sub-space. This will be defined by a system of Pfaffian relations in Plückerian co-ordinates of the mth dimension, like (32), and will determine an M sub-space at each point of the original n space. We assume that m is less than or equal to $M - 1$. The problem of finding geodesic Riemannian m-dimensional sub-spaces in the M dimensional Pfaffian may now be formulated, and solved by the method of Lagrangian multipliers. That is, we take as the integrand of the integral to be minimized:

$$L = [\sqrt{g_{i_1\cdots i_m, j_1\cdots j_m} a^{i_1\cdots i_m} a^{j_1\cdots j_m}} + S_p {}^{p}A_{i_1\cdots i_m} a^{i_1\cdots i_m}] du^1 \cdots du^m. \tag{35}$$

The S_p are variable multipliers. The conditions for a minimum are then given by (28), with the L therein defined by (35). For properly restricted boundaries, there will be, in general, a geodesic m-space. In particular, if there is an m-space which is geodesic for the Riemannian n space, passes through the given boundary and lies in the Pfaffian space it solves the present problem.

The dual of a problem of finding the m-space which minimizes the integral of m volume, and lies in a Pfaffian system which

restricts the m-space at each point to lie in an M space is a problem concerning an m space, but along which the elements of $m' = (n - m)$ space minimize the integral of m' volume and lie in a Pfaffian system which restricts the m'-space at each point to lie in an $M' = (n - M)$ space. The boundary conditions are the same.

Thus, for example, in three space we have the problem of finding the curve of minimum length in a single Pfaffian. Here the one dimensional element lies in a two space, and the curve sought joins fixed end points. That is, the end elements are among the vectors in the Pfaffian at these points. The dual problem again concerns a curve, but thought of as formed from the perpendicular two dimensional elements. Since the one dimensional elements of the original problem had to lie in a planar element, the two dimensional elements of the dual problem must contain the normal vector. The end elements are at the same points as before. Thus the solution of the dual problem is the same curve as that which solves the original problem, the duality merely applying to the elements at its points.

In the general case, each of the problems has as boundary conditions a closed $m - 1$ dimensional space, and the points which make up the configuration solving the problem fill up an m dimensional figure bounded by this $m - 1$ space. However, in the first case we form at each point an m volume in the given M sub-space, while in the second we form the dual $n - m$ volume containing the dual $n - M$ space. The integral and minimization applies to these volumes.

For the initial problem, we take m less than M, as otherwise there is no free parameter, and so no minimum problem.

5. *Expressions for Arc Length.* We may obtain alternative expressions for the arc length in our Riemannian space by introducing at each point an orthogonal entuple, defined by:

$$dy^i = p_i{}^j dx^i. \tag{36}$$

We then have:

$$ds^2 = (dy^i)^2 = (p_i{}^j dx^i)^2, \tag{37}$$

a sum of squares. This enables us to simplify the arc length for elements in a Pfaffian sub-space, since we may take $n - M$ of the vectors dy^i in the dual sub-space, so that the sum (37) will only contain M squares for elements in the Pfaffian sub-space. In terms of the dual co-ordinates, we have:

$$dS^2 = (P_{i_2 \cdots i_n}{}^i dx^{i_2 \cdots i_n})^2, \tag{38}$$

so that it reduces to M squares, also. Similar remarks apply to the elements:

$$ds_m{}^2 = (p_{i_1 \cdots i_m}{}^i dx^{i_1 \cdots i_m})^2, \tag{39}$$

and the dual

$$dS_{n-m}{}^2 = (P_{i_{m+1} \cdots i_n}{}^i dx^{i_{m+1} \cdots i_n})^2. \tag{40}$$

ON THE TIME ADMITTANCES OF TRANSMISSION NETWORKS

By John L. Barnes and Clarence T. Prendergast

Abstract

A method for determining the driving-point time admittances [1] of transmission networks is developed in this paper. The method employs network transformations. The driving-point time admittances of six types of wave-filter sections, and the transfer time admittances of six corrective networks are obtained by the method.

The transfer time admittances of two wave-filter half sections and of three other corrective networks are secured by operational methods.

Foreword

This paper presents a part of the Master's Thesis submitted by the authors at the Massachusetts Institute of Technology in May 1929. For advice and guidance in preparing the paper the authors wish to thank Professor V. Bush, Mr. J. R. Carson, and Dr. O. J. Zobel.

Introduction

For previous work done in the field of the present article the reader is referred to the paper (7) [2] of Carson and Zobel.

The importance and use of the driving-point [3] and transfer time admittances of an electrical network are brought out in (7, 19, 32, 35). Stated briefly, these two characteristic functions of a network are of high importance in precise studies of distortion

[1] *Time admittance* refers to the function of time called *indicial admittance* by J. R. Carson.

[2] Numbers given in this manner refer to the bibliography to be found at the end of the paper.

[3] See Appendix A.

and interference in electrical communication systems and of switching and lightning protection in power transmission systems.

The wave filters and corrective networks to be discussed here are types used in electrical communication systems. The methods employed in their treatment, however, are applicable to a more general class of networks which, of course, includes the types to be discussed. This class is characterized by the assumptions that the networks may be treated at all frequencies as circuits which are composed of lumped passive elements whose values are independent of time and of all conditions in the circuits. For cases in which circuit connections are varied at an assigned value of time, initial equilibrium is assumed. Effects due to the wires connecting the elements are assumed to be negligible.

1. A Network-Transformation Method

1.1 *Development of the Method*

A knowledge of the driving-point time admittances of transmission networks is useful in studies involving terminal conditions, as, for example, in the study of paralleling such networks.

The network-transformation (parallel-arm-expansion) method for determining the form of the driving-point time admittance of transmission networks which is now to be considered may be compared with the well-known Heaviside Expansion (see Appendix *D* and (1)). In using the Heaviside Expansion to determine the driving-point time admittance of a network, the network time admittance is expressed as a sum of simpler time admittances which are computed from the driving-point frequency [4] admittance of the network. In applying the network expansion to the same problem, the network is transformed (expanded) into a network having only parallel arms whose time admittances are already known or can be more easily computed. The network

[4] This adjective is used to distinguish between what is usually called the *admittance* of a network and its time admittance.

time admittance is then the sum of these simpler time admittances.

Extending this comparison to the more general case: the solution of the operational equation

$$A(t) = \frac{1}{Z(p)}\mathbf{1} = Y(p)\mathbf{1}$$

is often effected by an expansion of the right-hand side in the form

$$Y(p)\mathbf{1} = Y_1(p)\mathbf{1} + Y_2(p)\mathbf{1} + \cdots + Y_n(p)\mathbf{1},$$

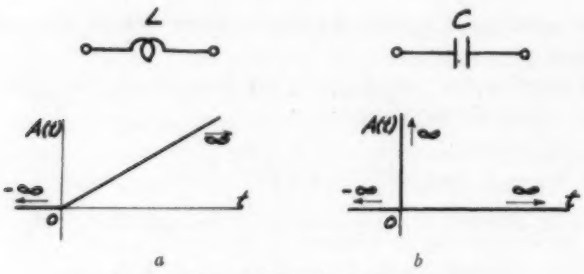
provided each component operational equation

$$h_m = Y_m(p)\mathbf{1}$$

is solvable. This expansion is usually effected analytically without explicit regard to the physical structure corresponding to $Y(p)$. With the present method, however, it is effected by an actual transformation of the network into an electrically equivalent physical structure.

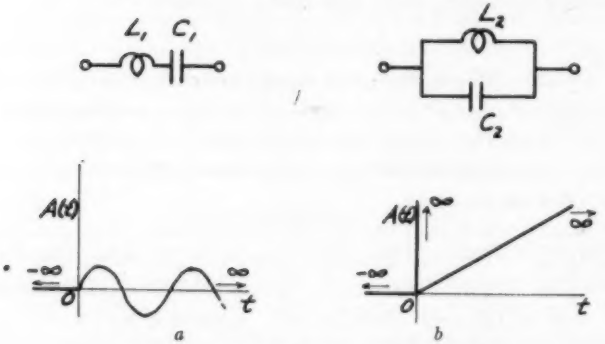
The network-transformation method is based on the following two facts. First, for networks composed only of parallel arms having no mutual elements, the network admittance is equal to the sum of the arm admittances. Second, most useful transmission networks can be transformed into equivalent networks having only parallel arms which are of this type, and which are simpler than the original network.

As shown in (15) and (38) transmission networks (for which our original assumptions hold) may be grouped according to the irreducible number of elements which they contain. Since ideally the elements of which a wave-filter section is composed are non-dissipative, these elements and the particular combinations of them most used in wave filters will be considered first. The combinations are the ones called by Zobel (6, 32) "inverse networks of impedance product R^2, or more concisely, inverse networks." The wave filters which are composed of inverse networks are of the constant-k type or of types derived from this (38, 44). In Figs. 1 to 5′ are shown all possible two terminal



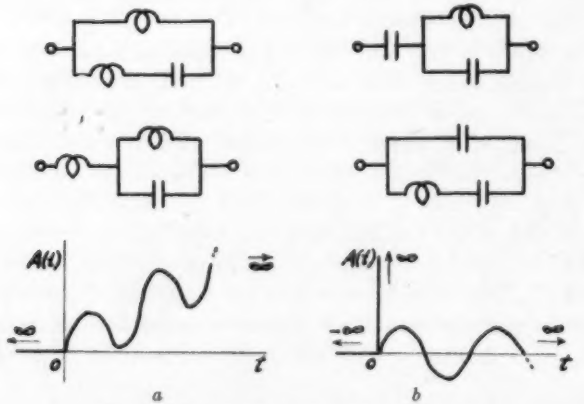
a b

FIG. 1. Inverse Networks—One Element.



a b

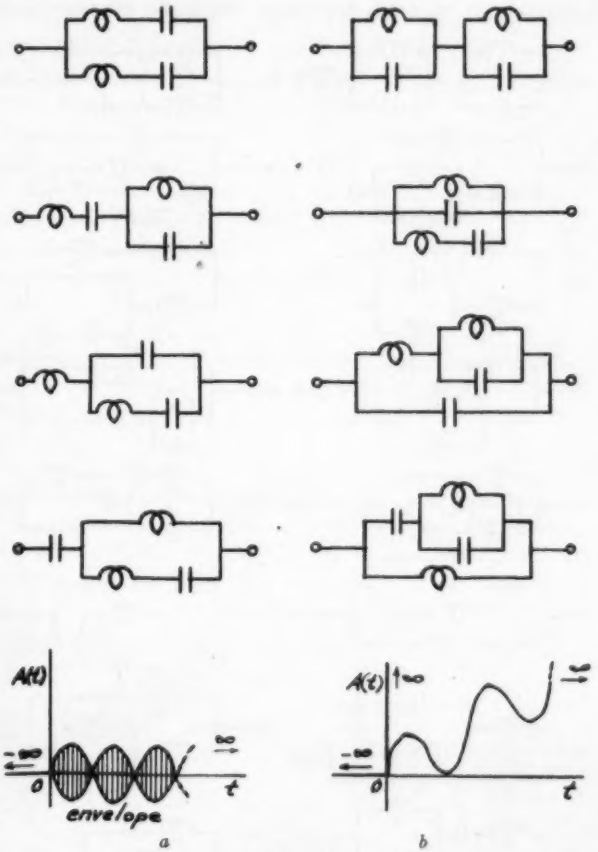
FIG. 2. Inverse Networks—Two Element.



a b

FIG. 3. Inverse Networks—Three Element.

network configurations in which appear an irreducible number of non-dissipative elements ranging in number from one to five (38).



a b

FIG. 4. Inverse Networks—Four Element.

The networks in part *a* of each figure are all potentially equivalent.[5] The same is true of those in part *b*. Also, every net-

[5] "Potentially equivalent" means "can be made equivalent by the proper assignment of element values" (38).

work in part *a* is potentially inverse to every network in part *b*. Combinations of more than five non-dissipative elements are not given, but may be derived in the same manner as those

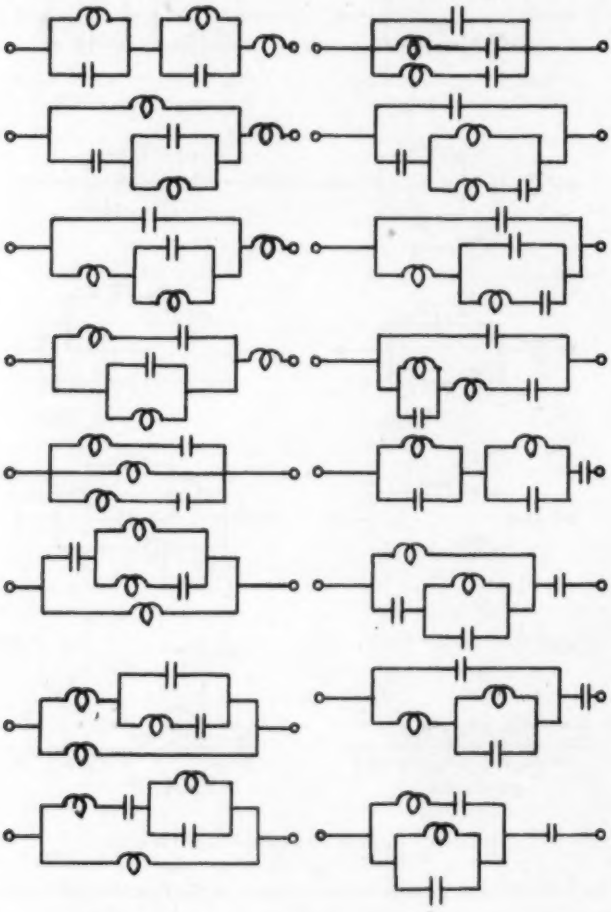


FIG. 5. Inverse Networks—Five Element.

shown (see (38)). From Figs. 2*b* to 5′ it may be observed that in each set of potentially equivalent networks there is one which is composed only of parallel arms, and that each of these parallel

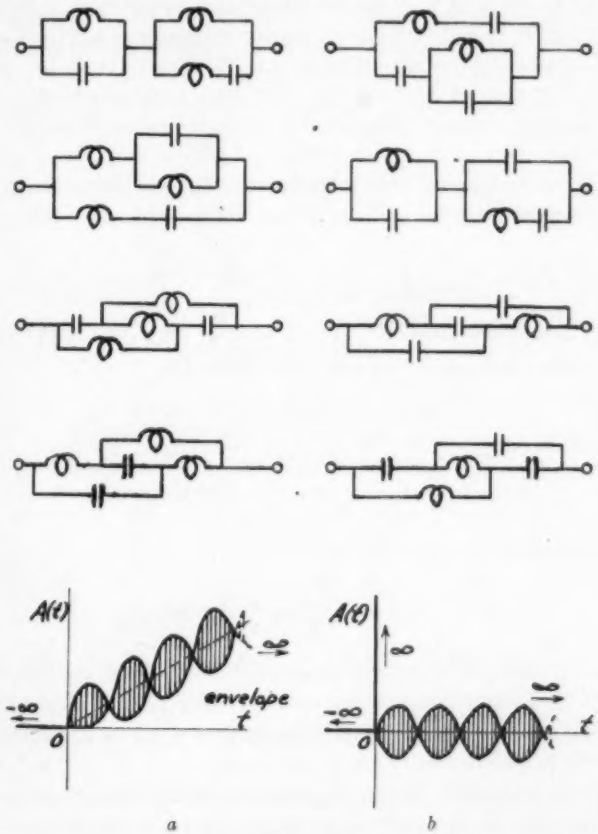


FIG. 5′. Inverse Networks—Five Element.

arms consists of one of the three networks shown in Figs. 1*a*, 1*b*, and 2*a*. Hence, it follows from the statement in the beginning of

this section that from the driving-point time admittances of an inductance, a capacitance, and an inductance and capacitance in series, the driving-point time admittances of all the other networks shown in Figs. 2*b* to 5′ can be synthesized by addition. By the same method the driving-point time admittances of any networks having non-dissipative elements can be synthesized, for such networks can be transformed (10) into equivalent networks having only parallel arms of the types shown in Figs. 1*a*, 1*b*, and 2*a*.

The driving-point time admittances of these three arms are obtained operationally (35) as follows. For inductance (Fig. 1*a*),

$$Y(p)\mathbf{1} = \frac{1}{Z(p)}\mathbf{1} = \frac{1}{pL}\mathbf{1}, \qquad \text{so} \qquad A(t) = \frac{t}{L}\mathbf{1}.$$

For capacitance (Fig. 1*b*), $Y(p)\mathbf{1} = pC\mathbf{1}$, so $A(t) = pC\mathbf{1}$. For inductance and capacitance in series (Fig. 2*a*),

$$Y(p)\mathbf{1} = \frac{1}{pL_1 + \dfrac{1}{pC_1}}\mathbf{1} = \sqrt{\frac{C_1}{L_1}}\,\frac{\dfrac{1}{\sqrt{L_1C_1}}\,p}{p^2 + \dfrac{1}{L_1C_1}}\mathbf{1},$$

so by formula 4, page 380 (35),

$$A(t) = \sqrt{\frac{C_1}{L_1}}\sin\frac{t}{\sqrt{L_1C_1}}\mathbf{1}.$$

As an example of the synthesis mentioned above, the driving-point time admittance of the networks shown in Fig. 3*a* is the sum of the driving-point time admittances of an inductance and a series-resonant arm (Figs. 1*a* and 2*a*).

It will be noticed that in the synthesis of the time admittances of the networks shown in Figs. 4*a*, 5′*a*, and 5′*b*, the sum of two sinusoidal waves of different frequencies is involved. In these and similar cases which follow use is made of the trigonometric identity,

$$\sin a + \sin b \equiv 2 \sin \tfrac{1}{2}(a + b) \cos \tfrac{1}{2}(a - b),$$

to obtain a "modulated" wave whose envelope is shown in the figures.

Next the dissipative networks corresponding to those shown in Figs. 1 to 5′ will be considered. In Figs. 6 to 11 are shown a resistance network together with dissipative networks corresponding to one in each of the groups of potentially equivalent networks in Figs. 1 to 5′, and their driving-point time admittances.

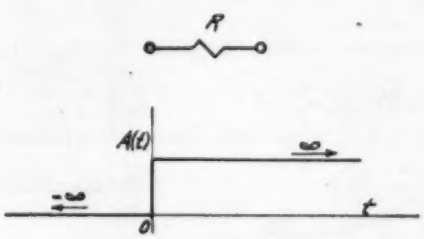


FIG. 6. Combinations of R, L, and C.

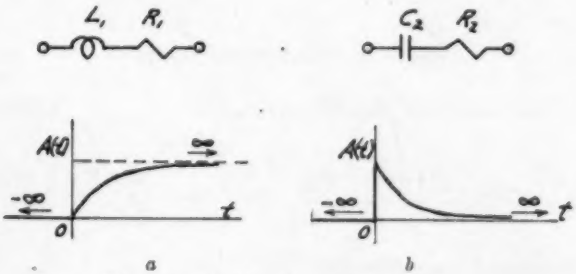


FIG. 7. Combinations of R, L, and C.

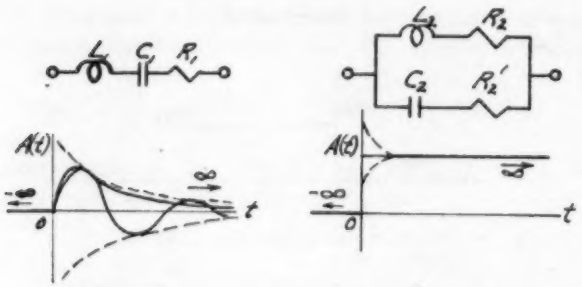
As in the case of the non-dissipative networks, from the driving-point time admittances of the arms in Figs. 7*a*, 7*b*, and 8*a*, we may synthesize the driving-point time admittances of all the other networks in Figs. 7 to 11 and their equivalent networks. The driving-point time admittances of the arms shown in Figs. 7*a*, 7*b*, and 8*a*, are obtained operationally (35) as follows:

For inductance and resistance (Fig. 7*a*),

$$Y(p)\mathbf{1} = \frac{1}{R_1 + pL_1}\mathbf{1} = \frac{1}{L_1}\frac{1}{\frac{R_1}{L_1} + p}\mathbf{1}.$$

By formula 2, page 380 (35),

$$A(t) = \frac{1}{R_1}(1 - \epsilon^{-(R_1/L_1)t})\mathbf{1}.$$



a *b*

FIG. 8. Combinations of R, L, and C.

For capacitance and resistance (Fig. 7*b*),

$$Y(p)\mathbf{1} = \frac{1}{R_1 + \frac{1}{pC_1}}\mathbf{1} = \frac{1}{R_1}\frac{p}{p + \frac{1}{R_1C_1}}\mathbf{1}.$$

By formula 3, page 380 (35),

$$A(t) = \frac{1}{R_1}\epsilon^{-(t/R_1C_1)}\mathbf{1}.$$

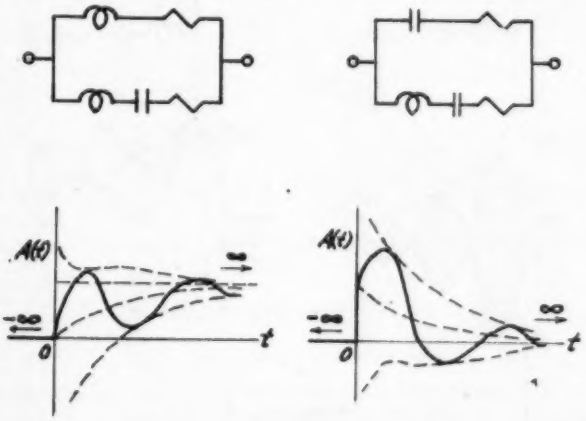
For inductance, capacitance, and resistance (Fig. 8*a*),

$$Y(p)\mathbf{1} = \frac{1}{R_1 + pL_1 + \frac{1}{pC_1}}\mathbf{1} = \frac{p}{p^2L_1 + pR + \frac{1}{C_1}}\mathbf{1}.$$

By formula 17, page 381 (35),

$$A(t) = \frac{e^{-(R_1 t/2L_1)}}{L_1\sqrt{\frac{1}{L_1C_1} - \frac{R_1^2}{4L_1^2}}} \sin \sqrt{\frac{1}{L_1C_1} - \frac{R_1^2}{4L_1^2}}\, t.$$

From the latter equation it may be seen (5) that we have a damped oscillation (see Fig. 8*a*) only when $R_1 < 2\sqrt{\frac{L_1}{C_1}}$. For $R_1 > 2\sqrt{\frac{L_1}{C_1}}$, $A(t)$ may be expressed as the product of a constant, an exponential damping factor, and a hyperbolic sine; hence its graph does not cross the time axis (see Fig. 8*a*). For $R_1 = 2\sqrt{\frac{L_1}{C_1}}$,



a *b*

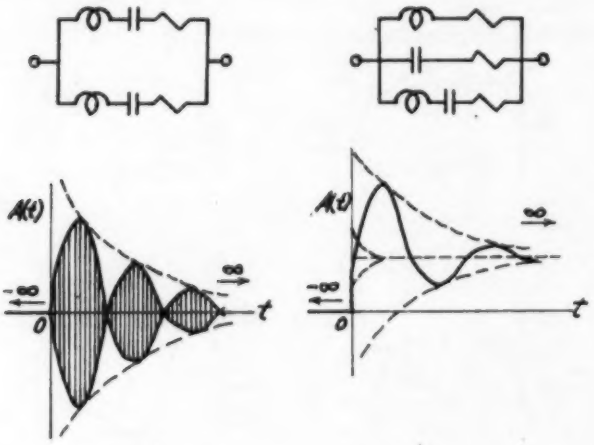
FIG. 9. Combinations of R, L, and C.

$A(t)$ may be expressed as the product of a constant, t (time), and an exponential damping factor; hence its graph does not cross the time axis either.

As an example of the above mentioned synthesis, the driving-point time admittance of the network shown in Fig. 10*a* is the

sum of two driving-point time admittances of the form shown in Fig. 8*a*.

The envelope in Fig. 10*a* was obtained under the assumption that $R < 2\sqrt{\frac{L}{C}}$ in each arm of the network. This assumption was also made in drawing the time admittance of the other figures in which the arm of Fig. 8*a* appears. The corresponding curves for $R \geqq 2\sqrt{\frac{L}{C}}$ in one or more arms may be easily constructed by substituting a non-oscillatory curve (see Fig. 8*a*).



a *b*

FIG. 10. Combinations of R, L, and C.

From Fig. 6 it may be seen that a resistance arm is distortionless for a unit wave.[6] Hence a resistance arm is distortionless for all waves (32).

From Figs. 7*a*, 7*b*, and 8*a*, it may be seen that among the effects produced by dissipation upon the driving-point time admittances of these networks are the reduction of infinite

[6] A *unit wave* is sometimes called a *unit function*.

pulses to finite damped admittances, and the reduction of finite admittances by an exponential damping factor. Also, in the case of sufficiently high dissipation, oscillatory time admittances are changed into non-oscillatory surges. By employing Figs. 6 to 11 it is possible to determine the form of the driving-point time admittances of many non-dissipative networks having dissipative terminal impedances, and also of many dissipative networks.

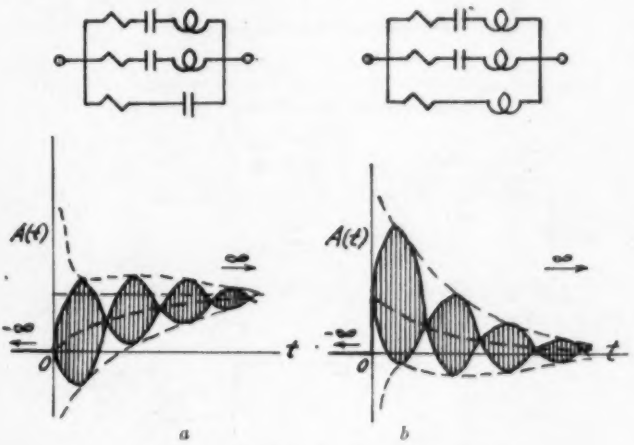


FIG. 11. Combinations of R, L, and C.

1.2 *Applications of the Method*

Several examples will be given to illustrate the application of this method for determining the time admittances of transmission networks.

In the first set of illustrations, the driving-point time admittance is determined for several low-, high-, and band-pass symmetrical-T wave-filter sections having $R = k$ terminations (38).

Referring to Fig. 12, it will be seen that a symmetrical-T wave-filter section having $R = k$ terminations may be transformed, as shown, by employing the transformations given at the

top of page 45 of (6). Replacing the small letters given in the reference by capitals, we have in this case, $A = 1$, $B = 2$, $C = 4$, and $D = 2$.

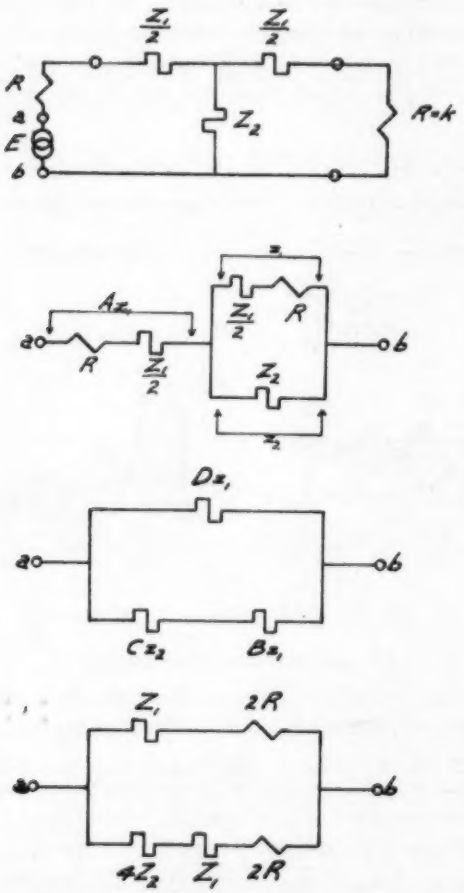


FIG. 12. Transformation of an $R = k$-Terminated Symmetrical-T Wave-Filter Section.

By substituting the values of the elements of the constant-k low-pass wave-filter section (Fig. 13a) in the transformed network shown in Fig. 12, the network shown in Fig. 13b is obtained.

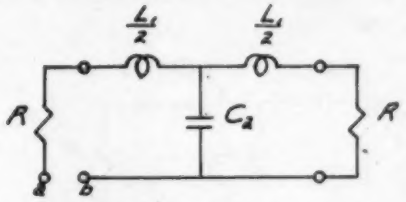


FIG. 13a. Transformation of an $R=k$-Terminated Symmetrical-T Constant-k Low-Pass Wave-Filter Section.

$$Z_1 = pL_1\,,\quad Z_2 = \frac{1}{pC_2}$$

$$L_1 = \frac{2R}{\omega_c}\,,\quad C_2 = \frac{2}{R\omega_c}$$

$$\omega_c = \frac{2}{\sqrt{L_1 C_2}}\,,\quad R = \sqrt{\frac{L_1}{C_2}}$$

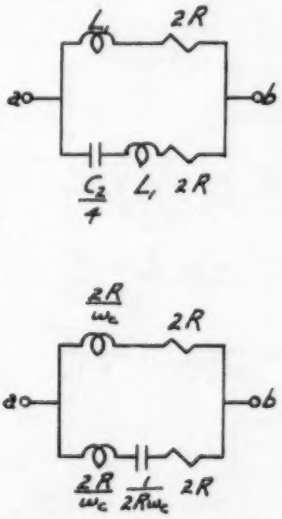


FIG. 13b.

This network is of the type shown in Fig. 9a. Thus the driving-point time admittance of this wave-filter section is of the form shown in Fig. 9a. In Appendix D this solution is checked

quantitatively by using the Heaviside Expansion Theorem (35).

A similar substitution of the element values of a constant-k high-pass wave-filter section (Fig. 14a) gives the network shown in Fig. 14b. The driving-point time admittance of this network is shown in Fig. 9b.

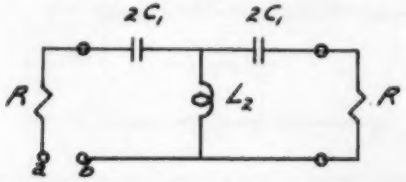
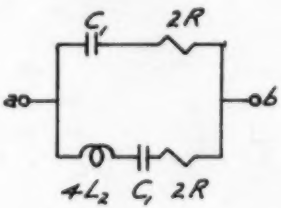


FIG. 14a. Transformation of an $R=k$-Terminated Symmetrical-T Constant-k High-Pass Wave-Filter Section.



$$Z_1 = \frac{1}{pC_1}, \quad Z_2 = pL_2$$

$$C_1 = \frac{1}{2\omega_c R}, \quad L_2 = \frac{R}{2\omega_c}$$

$$\omega_c = \frac{1}{2\sqrt{L_2 C_1}}, \quad R = \sqrt{\frac{L_2}{C_1}}$$

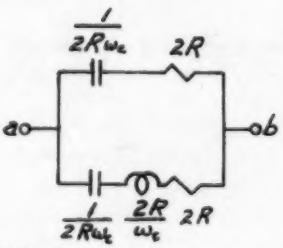


FIG. 14b.

Using the same procedure for the series M-type low- and high-pass wave-filter sections (Figs. 15a and 16a) the networks shown in Figs. 15b and 16b respectively are obtained. So the driving-

point time admittances of these sections have the forms shown in Figs. 9*a* and 9*b* respectively. It may be noted that the form of the driving-point time admittances of the series *M*-type sec-

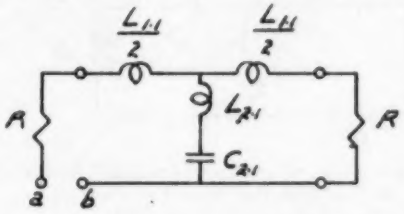


FIG. 15*a*. Transformation of an $R=k$-Terminated Symmetrical-T Series M-Type Low-Pass Wave-Filter Section.

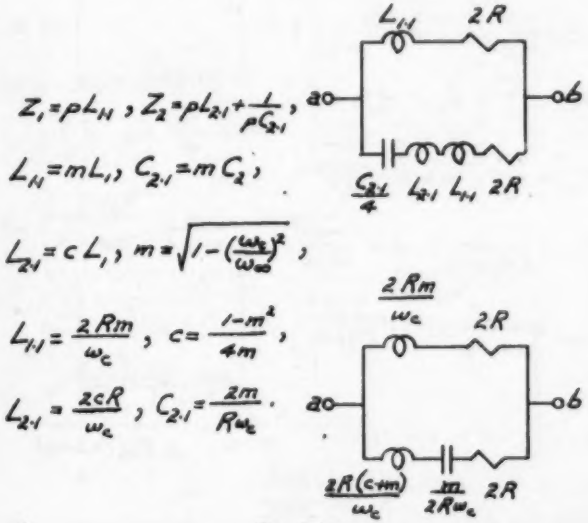


FIG. 15*b*.

tions is the same as that of their prototypes (see Figs. 9*a* and 9*b*). However, for the *M*-type sections, the frequency of oscillation, decrement coefficient, and amplitude factor of the time admit-

tance are functions of the parameter m, when $t > 0$ (see Appendix E).

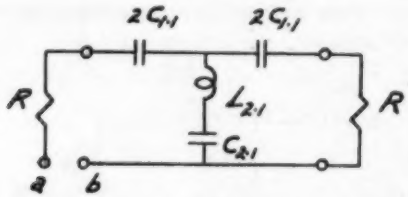


FIG. 16a. Transformation of an $R=k$-Terminated Symmetrical-T Series M-Type High-Pass Wave-Filter Section.

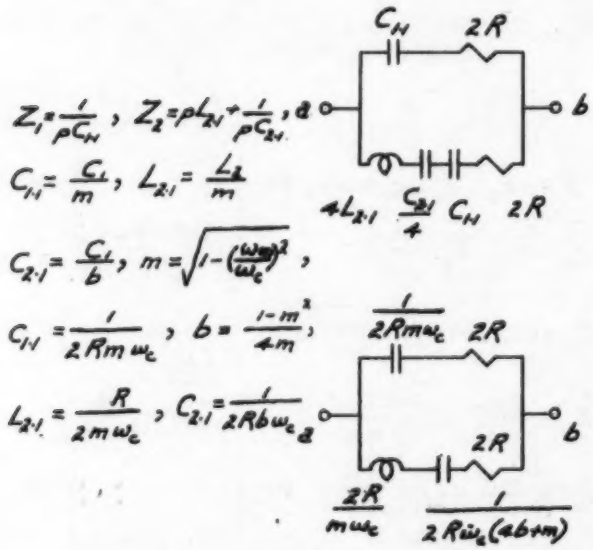


FIG. 16b.

Again, for $R=k$-terminated band-pass wave-filter sections of the $L_{1.1}C_{1.1}L_{2.1}$ and $L_{1.1}C_{1.1}C_{2.1}$ types (Figs. 17a and 18a) we obtain the networks shown in Figs. 17b and 18b respectively. So the driving-point time admittances of these sections have the form shown in Fig. 10a.

In Figs. 13 to 18 the network transformations have been carried out quantitatively, but the configurations of the networks are alone sufficient to determine the forms of the time admit-

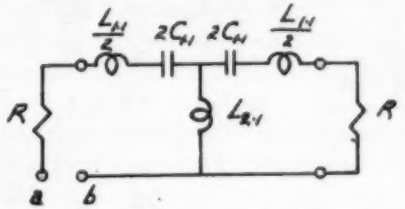


FIG. 17a. Transformation of a Symmetrical-T Type $L_{1\cdot1}C_{1\cdot1}L_{2\cdot1}$ Band-Pass Wave-Filter Section.

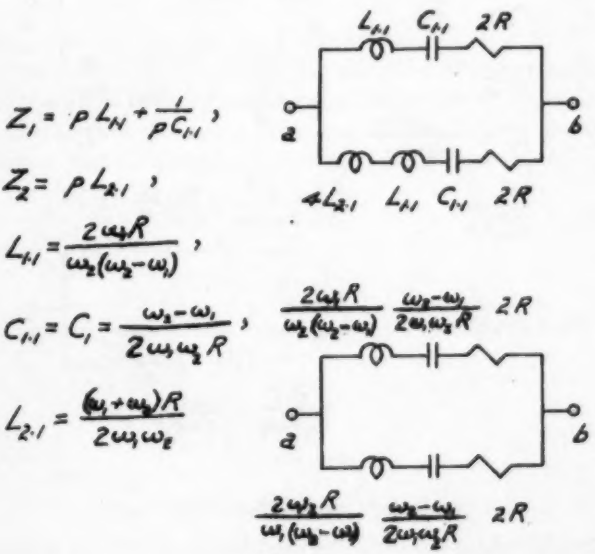


FIG. 17b.

tances. Thus when only the forms are required, the network transformations constitute a simple and short procedure for obtaining the results.

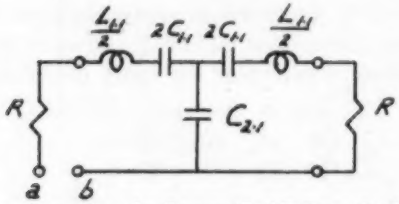


FIG. 18*a*. Transformation of a Symmetrical-*T* Type $L_{1,1}C_{1,1}C_{2,1}$ Band-Pass Wave-Filter Section.

$$Z_1 = pL_{1,1} + \frac{1}{pC_{1,1}},$$

$$Z_2 = \frac{1}{pC_{2,1}},$$

$$L_{1,1} = L_1 = \frac{2R}{\omega_2 - \omega_1},$$

$$C_{1,1} = \frac{\omega_2 - \omega_1}{2\omega_1^2 R},$$

$$C_{2,1} = \frac{2}{(\omega_2 + \omega_1)R}$$

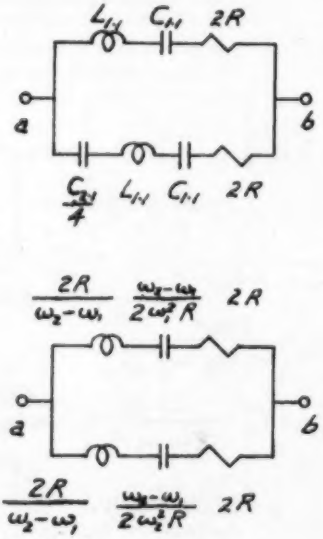


FIG. 18*b*.

1.3 *Obtaining Transfer Time Admittances*

Next the network-transformation method will be used to determine the form of the transfer time admittances of single sections of certain corrective networks.

It has been pointed out (32) that one of the fundamental methods used in distortion correction is to alter the system

transmission characteristics by the insertion of one or more sections of constant-resistance recurrent networks. Thus, by design, the driving-point time admittances of these corrective networks have the form shown in Fig. 6. The transfer time admittances of the corrective-network sections shown in Fig. 19 will next be determined. Zobel has already treated a special case in (32).

The propagation constants of single sections of the networks in Fig. 19 have been shown by Zobel (32) to be all equal, and to have the value

$$P = ln\left(1 + \frac{Z_{11}}{R}\right). \tag{1}$$

When one of these sections, for example c in Fig. 19, is terminated by resistances of value R, the transfer time admittance may be determined as follows. Let E be an electromotive force applied to the section through a resistance R. Let I_0 be the input current, and I_1 the output current. Then we know from (32 or 18) that

$$I_1 = I_0 e^{-P}, \tag{2}$$

and

$$I_0 = \frac{E}{2R}. \tag{3}$$

Substituting values from Eqs. (1) and (3) in (2), it follows that

$$I_1 = \frac{E}{2R}\frac{1}{1 + \frac{Z_{11}}{R}} = \frac{E}{2}\frac{1}{R + Z_{11}}. \tag{4}$$

So,

$$Y_1(p)\mathbf{1} = \frac{1}{2}\frac{1}{R + Z_{11}}\mathbf{1} = A_1(t). \tag{5}$$

It follows from Eq. (5) that the transfer time admittances of the networks shown in Fig. 19, when R-terminated, have the form of the driving-point time admittance of the series combination R and Z_{11}. As an example, if $Z_{11} = pL_{11}$, the transfer time admittance of any one of these sections has the form shown in Fig. 7*a*.

Or, if $Z_{11} = \dfrac{1}{pC_{11}}$, the form of the transfer time admittance is shown in Fig. 7*b*.

The foregoing examples illustrate the procedure used in deter-

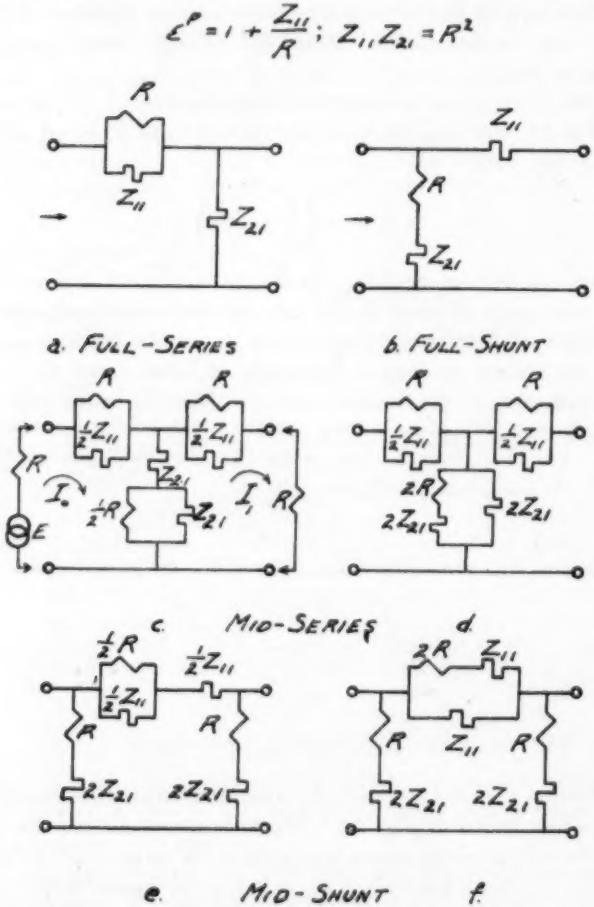


FIG. 19. Ladder-Type Constant Resistance Sections.

mining the form of the time admittance of a transmission network.

It may now be remarked that there are two terminal networks satisfying our original assumptions, which cannot be transformed so as to have only parallel branches composed of series combinations of resistance, self inductance, and capacitance (12). An example of such a network is an $R=k$-terminated section of a constant-k band-pass wave filter. The transformation given in Fig. 12, however, reduces this network to a form which simplifies the direct operational and Heaviside Expansion Theorem solutions.[7] Thus, in this example, the network transformation method for finding the time admittances fails to give a solution of the problem, but does simplify it.

Although the method has been used above only to obtain the form of the time admittances, by employing the quantitative network transformations it can be used as a quantitative method for simplifying the computation of network time admittances.

2. Determination of the Transfer Time Admittances of Certain Wave-Filter Half Sections and Corrective Networks

2.1 *Image-Terminated Half-Sections of Constant-k Low- and High-Pass Wave Filters*

In the design of composite wave filters it is desirable to terminate half sections by image frequency impedances (38). Such terminations are, however, "infinite" terminations and cannot be synthesized from a few circuit elements (R, L, and C).[8]

The transfer time admittance of an image-terminated constant-k low-pass wave-filter half section (see Fig. 20) may be computed operationally as follows. The parameters of a constant-k low-pass wave filter are (7):

$$L_1 = \frac{2k}{\omega_c}, \qquad C_2 = \frac{2}{\omega_c k}, \qquad k = \sqrt{Z_1 Z_2}, \qquad \omega_c = \frac{2}{\sqrt{L_1 C_2}},$$

[7] In each case a 4th degree polynomial replaces one of 6th degree.

[8] Such terminations can, nevertheless, be approximated in various ways by combinations of a few circuit elements (41, 42, 44).

so

$$\frac{Z_1}{4Z_2} = \left(\frac{p}{\omega_c}\right)^2.$$

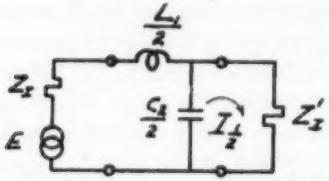


FIG. 20. Image-Terminated Constant-k Low-Pass Wave-Filter Half Section.

From Appendix C,

$$I_{\frac{1}{2}} = \frac{E}{2\sqrt{Z_1 Z_2}} \frac{\sqrt{1 + \frac{Z_1}{4Z_2}} - \sqrt{\frac{Z_1}{4Z_2}}}{\sqrt{1 + \frac{Z_1}{4Z_2}}}.$$

So, substituting the values of the parameters,

$$Y_{\frac{1}{2}}(p)\mathbf{1} = \frac{1}{2k} \frac{\sqrt{1 + \left(\frac{p}{\omega_c}\right)^2} - \frac{p}{\omega_c}}{\sqrt{1 + \left(\frac{p}{\omega_c}\right)^2}} \mathbf{1} = \frac{1}{2k} \frac{\sqrt{p^2 + \omega_c^2} - p}{\sqrt{p^2 + \omega_c^2}} \mathbf{1}.$$

Then,

$$A_{\frac{1}{2}}(t) = \frac{\omega_c}{2k} \frac{1}{p} \frac{p}{\sqrt{p^2 + \omega_c^2}} \frac{\sqrt{p^2 + \omega_c^2} - p}{\omega_c} \mathbf{1}.$$

Applying formula 48, page 382 (35),

$$A_{\frac{1}{2}}(t) = \frac{\omega_c}{2k} \int_0^t J_1(\omega_c \lambda) d\lambda \mathbf{1} = \frac{1}{2k} \int_0^{\omega_c t} J_1(x) dx \mathbf{1}.$$

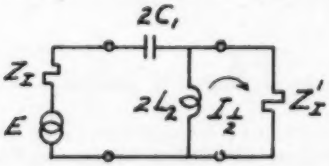


FIG. 21. Image-Terminated Constant-k High-Pass Wave-Filter Half Section.

The transfer time admittance of an image-terminated constant-k high-pass wave-filter half section (see Fig. 21) may be computed in the same way. The parameters of a constant-k high-pass wave filter are (7):

$$C_1 = \frac{1}{2\omega_c k}, \qquad L_2 = \frac{k}{2\omega_c}, \qquad k = \sqrt{Z_1 Z_2}, \qquad \omega_c = \frac{1}{2\sqrt{L_2 C_1}},$$

so

$$\frac{Z_1}{4Z_2} = \left(\frac{\omega_c}{p}\right)^2.$$

From Appendix C,

$$I_1 = \frac{E}{2\sqrt{Z_1 Z_2}} \frac{\sqrt{1 + \frac{Z_1}{4Z_2}} - \sqrt{\frac{Z_1}{4Z_2}}}{\sqrt{1 + \frac{Z_1}{4Z_2}}}.$$

Substituting the values of the parameters,

$$Y_1(p)\mathbf{1} = \frac{1}{2k} \frac{\sqrt{1 + \left(\frac{\omega_c}{p}\right)^2} - \frac{\omega_c}{p}}{\sqrt{1 + \left(\frac{\omega_c}{p}\right)^2}} \mathbf{1} = \frac{1}{2k}\left(1 - \frac{\omega_c}{\sqrt{p^2 + \omega_c^2}}\right)\mathbf{1}.$$

So,

$$A_1(t) = \frac{1}{2k}\mathbf{1} - \frac{\omega_c}{2k}\frac{1}{p}\frac{p}{\sqrt{p^2 + \omega_c^2}}\mathbf{1}.$$

Applying formula 46, page 382 (35),

$$A_1(t) = \frac{1}{2k}\mathbf{1} - \frac{\omega_c}{2k}\int_0^t J_0(\omega_c \lambda) d\lambda \mathbf{1}$$

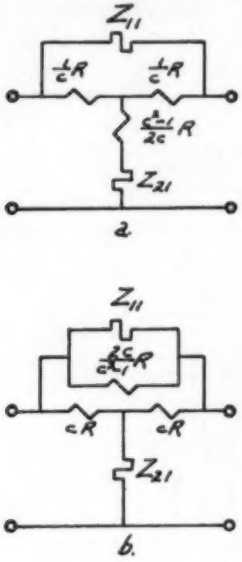
$$= \frac{1}{2k}\mathbf{1} - \frac{1}{2k}\int_0^{\omega_c t} J_0(x) dx \mathbf{1}.$$

The "infinite" terminations may be said to be responsible for the Bessel-function terms in these two transfer time admittances.

2.2 *Single Sections of Certain Bridged-T Constant-Resistance Corrective Networks*

Consider next the determination of the transfer time admittances of the bridged-T constant-resistance corrective-network

sections shown in Fig. 22. Zobel has shown (32) that the propagation constants of these sections are equal, and have the



$$c \geq 1$$

$$Z_{11}Z_{21} = R^2$$

$$\epsilon^P = \frac{1 + \frac{c+1}{2R}Z_{11}}{1 + \frac{c-1}{2R}Z_{11}}$$

c = an arbitrary real design constant.

FIG. 22. Certain Bridged-T Type Constant Resistance Sections.

value

$$P = ln\left(\frac{1 + \frac{c+1}{2R}Z_{11}}{1 - \frac{c-1}{2R}Z_{11}}\right) \tag{6}$$

When either of these sections is terminated by resistances of value R, its transfer time admittance may be determined as follows. Let E be an electromotive force applied to the section through a resistance R. Let I_0 be the input current and I_1 the output current. Then from (18),

$$I_1 = I_0 \epsilon^{-p}, \tag{2}$$

and

$$I_0 = \frac{E}{2R}. \tag{3}$$

Substituting from Eqs. (6) and (3) in (2),

$$I_1 = \frac{E}{2R} \frac{1 + \frac{c-1}{2R} Z_{11}}{1 + \frac{c+1}{2R} Z_{11}} \tag{7}$$

So,

$$Y_1(p)\mathbf{1} = \frac{1}{2R} \frac{1 + \frac{c-1}{2R} Z_{11}}{1 + \frac{c+1}{2R} Z_{11}} \mathbf{1} = A_1(t).$$

This expression may be written in the following form,

$$A_1(t) = \frac{1}{2R} \frac{2R}{c+1} \frac{1}{\frac{2R}{c+1} + Z_{11}} \mathbf{1} + \frac{1}{2R} \frac{c-1}{2R} \frac{Z_{11}}{1 + \frac{c+1}{2R} Z_{11}} \mathbf{1}. \tag{9}$$

For the case in which $Z_{11} = pL_{11}$,

$$A_1(t) = \frac{1}{2R} \frac{2R}{L_{11}(c+1)} \frac{1}{\frac{2R}{L_{11}(c+1)} + p} \mathbf{1} + \frac{1}{2R} \frac{c-1}{c+1} \frac{p}{\frac{2R}{L_{11}(c+1)} + p} \mathbf{1}.$$

Applying formulas (2) and (3) on page 380 (35),

$$A_1(t) = \frac{1}{2R}(1 - \epsilon^{-(2R/L_{11}(c+1))t})\mathbf{1} + \frac{1}{2R}\frac{c-1}{c+1}\epsilon^{-(2R/L_{11}(c+1))t}\mathbf{1}$$

$$= \frac{1}{2R}\left(1 - \frac{2}{c+1}\epsilon^{-(2R/L_{11}(c+1))t}\right)\mathbf{1}. \tag{10}$$

Notice that in this case the transfer time admittance has the same form as the driving-point time admittance of an inductance and resistance in series (see Fig. 7*a*).

For the case in which $Z_{11} = \dfrac{1}{pC_{11}}$,

$$A_1(t) = \frac{1}{2R}\left(\frac{p}{\dfrac{c+1}{2RC_{11}} + p}\mathbf{1} + \frac{c-1}{2RC_{11}}\frac{1}{\dfrac{c+1}{2RC_{11}} + p}\mathbf{1}\right).$$

Applying formulas (3) and (2) on page 380 (35),

$$A_1(t) = \frac{1}{2R}\left[\epsilon^{-(c+1/2RC_{11})t} + \frac{c-1}{c+1}(1 - \epsilon^{-(c+1/2RC_{11})t})\right]\mathbf{1}$$

$$= \frac{1}{2R}\left(\frac{c-1}{c+1} + \frac{2}{c+1}\epsilon^{-(c+1/2RC_{11})t}\right)\mathbf{1}. \tag{11}$$

Notice in this case that the transfer time admittance has the same form as the driving-point time admittance of a capacitance and resistance in series shunted by a resistance.

It may also be observed that if in Eq. (9) c is put equal to the limiting value 1, Eq. (5) is obtained. Hence for $c = 1$, the remarks made about Eq. (5) apply here too.

2.3 *The Iteratively Terminated Balanced Lattice-Type Corrective Network*

Mason in (34) obtained the time-response characteristic of the nth section of the iteratively terminated balanced lattice-type corrective network shown in Fig. 23 to a cisoidal wave (4) applied at time zero. He used a method which involves the expression of the network elements as limiting cases of a dis-

tortionless transmission line, a series expansion of the resulting expression for the response current, and a summation of n terms of the series. An operational solution of a nearly equivalent problem, that of determining the transfer time admittance of the nth section of this lattice network, will now be given.

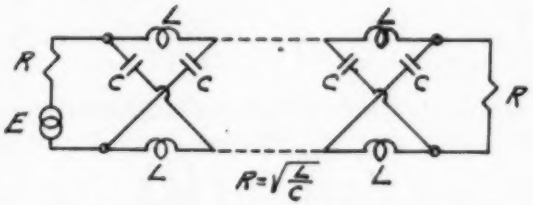


FIG. 23. Iteratively Terminated Balanced Lattice-Type Corrective Network.

From Appendix F we know that the current in the nth section of a lattice network such as that shown in Fig. 23 is

$$I_n = \frac{E}{2\sqrt{Z_1 Z_2}} \left(\frac{1 - \frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}{1 + \frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}} \right)^n. \tag{12}$$

The parameters of this network are (4),

$$\frac{1}{2} Z_1 = pL, \qquad 2Z_2 = \frac{1}{pC}, \qquad R = \sqrt{\frac{L}{C}},$$

so

$$Z_1 = 2pL, \qquad Z_2 = \frac{1}{2pC}, \qquad \text{and} \qquad \sqrt{\frac{Z_1}{Z_2}} = 2p\sqrt{LC}.$$

Substituting these values in Eq. (12),

$$I_n = \frac{E}{2R} \left(\frac{1 - p\sqrt{LC}}{1 + p\sqrt{LC}} \right)^n, \tag{13}$$

and the transfer time admittance is

$$A_n(t) = Y_n(p)1 = \frac{1}{2R} \left(\frac{1 - p\sqrt{LC}}{1 + p\sqrt{LC}} \right)^n 1. \tag{14}$$

For convenience, let $\sqrt{LC} = b$ in Eq. (13). Then,

$$I_n(t) = \frac{E}{2R}(-1)^n \left(\frac{bp-1}{bp+1}\right)^n 1. \tag{15}$$

Carson has shown that an operational equation of this type may be solved by the following method (19). Defining a function φ_n by

$$\varphi_n \equiv \left(\frac{p-1}{p+1}\right)^n, \tag{16}$$

and using Carson's Theorem VI (*B. S. T. Jl.*, Oct. 1925, p. 728), we obtain

$$I_n(t) = \frac{E}{2R}(-1)^n \varphi_n\left(\frac{t}{b}\right). \tag{17}$$

It is therefore first necessary to evaluate φ_n in Eq. (16). From the auxiliary operational expression

$$\sigma_n \equiv \left(\frac{p-2}{p}\right)^n = \left[p^n - 2\frac{n}{1!}p^{n-1} + 2^2\frac{n(n-1)}{2!}p^{n-2} + \cdots + (-1)^n 2^n\right]\frac{1}{p^n}$$

the explicit solution for σ_n is obtained by replacing $\frac{1}{p^n}$ by $\frac{t^n}{n!}$ and p^n by $\frac{d^n}{dt^n}$, so

$$\sigma_n = \left[\frac{d^n}{dt^n} - 2\frac{n}{1!}\frac{d^{n-1}}{dt^{n-1}} + \cdots + (-1)^n 2^n\right]\frac{t^n}{n!}$$

$$= 1 - \frac{n(2t)}{1!\,1!} + \frac{n(n-1)(2t)^2}{2!\,2!} + \cdots + (-1)^n\frac{(2t)^n}{n!}. \tag{18}$$

Letting

$$\sigma_n = \frac{1}{H(p)} \equiv \left(\frac{p-2}{p}\right)^n,$$

it follows that

$$\varphi_n = \left(\frac{p-1}{p+1}\right)^n \equiv \frac{1}{H(p+1)}$$

$$= \frac{p+1}{p}\cdot\frac{p}{p+1}\cdot\frac{1}{H(p+1)} = \left(1+\frac{1}{p}\right)\cdot\frac{p}{p+1}\cdot\frac{1}{H(p+1)}.$$

So, applying Theorem V in the above mentioned paper,

$$\varphi_n(t) = \sigma_n(t)\epsilon^{-t} + \int_0^t \sigma_n(t)\epsilon^{-t}dt. \tag{19}$$

Substituting the value of φ_n in Eq. (17),

$$I_n(t) = \frac{E}{2R}(-1)^n\left[\sigma_n\left(\frac{t}{\sqrt{LC}}\right)\epsilon^{-t/\sqrt{LC}} + \int_0^{t/\sqrt{LC}} \sigma_n(\tau)\epsilon^{-\tau}d\tau\right]\mathbf{1}. \tag{20}$$

where τ is a variable of integration. The transfer time admittance of the nth section is therefore given by

$$A_n(t) = \frac{(-1)^n}{2R}\left[\sigma_n\left(\frac{t}{\sqrt{LC}}\right)\epsilon^{-t/\sqrt{LC}} + \int_0^{t/\sqrt{LC}} \sigma_n(\tau)\epsilon^{-\tau}d\tau\right]\mathbf{1}, \tag{21}$$

where the value of σ_n is given by Eq. (18).

For the case in which $n = 1$, it follows from Eqs. (21) and (18) that

$$A_1(t) = \frac{-1}{2R}\left[\left(1 - \frac{2t}{\sqrt{LC}}\right)\epsilon^{-t/\sqrt{LC}} + \int_0^{t/\sqrt{LC}} (1 - 2\tau)\epsilon^{-\tau}d\tau\right]\mathbf{1}$$

$$= \frac{1}{2R}(1 - 2\epsilon^{-t/\sqrt{LC}})\,\mathbf{1}. \tag{22}$$

If only the solutions for the cases of $n = 1$ and $n = 2$ are wanted, it will be found that the direct operational method is shorter than the evaluation of the more general solution given above.

The current response of a single section of the network to a cisoidal wave (4) applied at time zero will now be secured by the direct operational method. The operational formula for such a wave is given by 3 on page 380 (35) as

$$\frac{p}{p+b}\,\mathbf{1} = \epsilon^{-bt}, \qquad \text{if} \qquad b = -j\omega. \tag{23}$$

The substitution of this operational expression for the unit wave

in Eq. (14) gives the cisoidal time admittance $\alpha_1(t)$. Letting $\frac{1}{\sqrt{LC}} = a$ and $\sqrt{LC} = b$,

$$\alpha_1(t) = \frac{1}{2R}\frac{a-p}{a+p}\frac{p}{p+b}\mathbf{1}$$

$$= \frac{a}{2R}\frac{p}{(p+a)(p+b)}\mathbf{1} - \frac{1}{2R}\frac{p^2}{(p+a)(p+b)}\mathbf{1}. \quad (24)$$

Applying formulas 12 and 13 on page 381 (35),

$$\alpha_1(t) = \frac{a}{2R}\frac{1}{a-b}(\epsilon^{-bt} - \epsilon^{-at})\mathbf{1} - \frac{1}{2R}\frac{1}{b-a}(b\epsilon^{-bt} - a\epsilon^{-at})\mathbf{1}$$

$$= \frac{1}{2R}\frac{(1 - j\omega\sqrt{LC})\epsilon^{j\omega t} - 2\epsilon^{-(t/\sqrt{LC})}}{1 + j\omega\sqrt{LC}}\mathbf{1}. \quad (25)$$

By taking the real part of Eq. (25) the response of the section to a cosine wave applied at time zero is obtained. Removing the imaginary terms from the denominator by multiplying the numerator and denominator by the complex conjugate of the denominator gives

$$\alpha_1(t) = \frac{1}{2R}\frac{(1 - \omega^2 LC - 2j\omega\sqrt{LC})\epsilon^{j\omega t} - (2 - 2j\omega\sqrt{LC})\epsilon^{-(t/\sqrt{LC})}}{1 + \omega^2 LC}\mathbf{1}. \quad (26)$$

Taking the real part of Eq. (26) gives the cosine time admittance $\alpha_{1r}(t)$.

$$\alpha_{1r}(t) = \frac{1}{2R}\frac{(1 - \omega^2 LC)\cos\omega t + 2\omega\sqrt{LC}\sin\omega t - 2\epsilon^{-(t/\sqrt{LC})}}{1 + \omega^2 LC}\mathbf{1}. \quad (27)$$

By substituting $\omega\sqrt{LC} = \tan\varphi$ in Eq. (27) and employing several trigonometric identities there results

$$\alpha_{1r}(t) = \frac{1}{2R}[\cos(\omega t - 2\varphi) - 2\epsilon^{-\frac{t}{\sqrt{LC}}}\cos^2\varphi]\mathbf{1}. \quad (28)$$

Eq. (28) checks with the result obtained by Mason (34) for this case; and, as indicated by him, reduces to the transfer time admittance, given by Eq. (22), when $\omega = 0$.

Conclusion

To the results secured by Carson and Zobel (7) this paper adds the driving-point and transfer time admittances of a number of transmission networks. Some of these results were obtained by a method based on network transformations. This method has been described and its use illustrated.

These time admittances can now be used by the designer of transmission systems in the selection of networks which satisfy the minimum distortion and interference requirements.

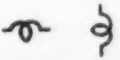
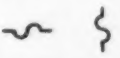
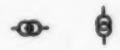
A more extensive study of the time admittances of transmission networks than has been made in the past will considerably help the study of the response of these networks to waves containing rapid changes and discontinuities. The field is rather important, for in it lie many of the major unsolved distortion and interference problems.

Appendix *A*

Terms and Symbols

The terms and symbols employed in this paper are either those used in well-known texts on the subjects treated, such as (18), (35), and (38), or are defined where they first appear.

In circuit diagrams the following symbols are used for the elements:

Name	*Symbol* *Mathematical*	*Diagramatic*
Resistance	R	
Inductance	L	
Capacitance	C	
Reactance	X	
Impedance	Z	
Electromotive Force	E	

APPENDIX *B*

Derivation of the Formula for the Current in the *n*th Section of an Iteratively-Terminated Recurrent Symmetrical Structure

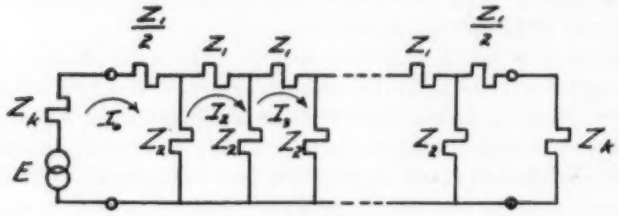


Fig. *B*1

In Fig. *B*1, Z_1 and Z_2 are the series and shunt arm impedances, and Z_k is the iterative impedance.

Let P denote the propagation constant per section of the structure, and I_n the current in the nth mesh of the structure. Then I_0 will denote the current entering the structure.

Now from (18) we know that

$$I_n = I_0 \epsilon^{-nP}, \tag{1}$$

$$I_0 = \frac{E}{2Z_k}, \tag{2}$$

and

$$Z_k = \sqrt{Z_1 Z_2}\sqrt{1 + \frac{Z_1}{4Z_2}}, \tag{3}$$

since the structure is mid-series terminated. Also,

$$P = ln\left(\frac{Z_1}{4Z_2} + 1 + \sqrt{\frac{Z_1}{Z_2} + \frac{Z_1^2}{4Z_2^2}}\right).$$

So by factoring,

$$P = ln\left(\sqrt{1 + \frac{Z_1}{4Z_2}} + \sqrt{\frac{Z_1}{4Z_2}}\right)^2. \tag{4}$$

Eq. (4) may also be written in the form

$$P = ln\left(\sqrt{1+\frac{Z_1}{4Z_2}} - \sqrt{\frac{Z_1}{4Z_2}}\right)^{-2}.$$

Hence

$$-nP = ln\left(\sqrt{1+\frac{Z_1}{4Z_2}} \pm \sqrt{\frac{Z_1}{4Z_2}}\right)^{\mp 2n}. \tag{5}$$

So

$$e^{-nP} = \left(\sqrt{1+\frac{Z_1}{4Z_2}} \mp \sqrt{\frac{Z_1}{4Z_2}}\right)^{\mp 2n}. \tag{6}$$

Substituting values from Eqs. (2), (3), and (6) in (1),

$$I_n = \frac{E}{2\sqrt{Z_1 Z_2}} \frac{\left(\sqrt{1+\frac{Z_1}{4Z_2}} \mp \sqrt{\frac{Z_1}{4Z_2}}\right)^{\mp 2n}}{\sqrt{1+\frac{Z_1}{4Z_2}}}. \tag{7}$$

So

$$Y_n(p) = \frac{\left(\sqrt{1+\frac{Z_1}{4Z_2}} \mp \sqrt{\frac{Z_1}{4Z_2}}\right)^{\mp 2n}}{2\sqrt{Z_1 Z_2}\sqrt{1+\frac{Z_1}{4Z_2}}}. \tag{8}$$

APPENDIX C

Derivation of the Formula for the Output Current of a Single Mid-Series Half Section with Image Terminations

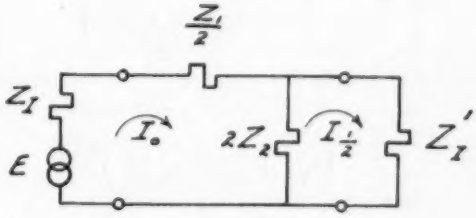


Fig. C1

In Fig. C1, Z_1 and Z_2 are the series and shunt impedances of a full section. Z_I and Z_I' are the mid-series and mid-shunt image impedances.

Let I_0 denote the input current, I_1 the output current, and θ the image transfer constant per section. Then $\frac{\theta}{2}$ will denote the image transfer constant for a half-section (18).

Now from (18) we know that

$$I_1 = I_0 \epsilon^{-(\theta/2)}, \tag{1}$$

$$I_0 = \frac{E}{2Z_I}, \tag{2}$$

$$Z_I = \sqrt{Z_1 Z_2}\sqrt{1 + \frac{Z_1}{4Z_2}}, \tag{3}$$

and

$$Z_I' = \frac{\sqrt{Z_1 Z_2}}{\sqrt{1 + \frac{Z_1}{4Z_2}}}. \tag{4}$$

Also

$$\frac{\theta}{2} = ln\left(\frac{Z_1}{2Z_2} + 1 + \sqrt{\frac{Z_1}{Z_2} + \frac{Z_1^2}{4Z_2^2}}\right)^{\frac{1}{2}}.$$

So by factoring as in Appendix B

$$\frac{\theta}{2} = ln\left(\sqrt{1 + \frac{Z_1}{4Z_2}} \mp \sqrt{\frac{Z_1}{4Z_2}}\right)^{\pm 1}. \tag{5}$$

Substituting values from Eqs. (2), (3), and (5) in (1),

$$I_1 = \frac{E}{2\sqrt{Z_1 Z_2}} \frac{\left(\sqrt{1 + \frac{Z_1}{4Z_2}} \mp \sqrt{\frac{Z_1}{4Z_2}}\right)^{\pm 1}}{\sqrt{1 + \frac{Z_1}{4Z_2}}}. \tag{6}$$

So

$$Y_1(p) = \frac{\left(\sqrt{1 + \frac{Z_1}{4Z_2}} \mp \sqrt{\frac{Z_1}{4Z_2}}\right)^{\pm 1}}{2\sqrt{Z_1 Z_2}\sqrt{1 + \frac{Z_1}{4Z_2}}}. \tag{7}$$

APPENDIX *D*

QUANTITATIVE DETERMINATION OF THE DRIVING-POINT TIME ADMITTANCE OF A SYMMETRICAL-*T* CONSTANT-*k* LOW-PASS WAVE-FILTER SECTION WITH $R = k$ TERMINATIONS

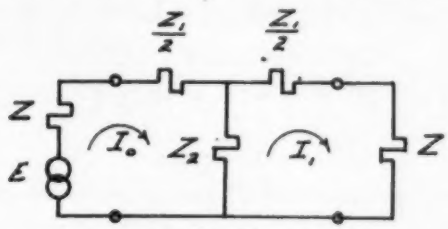


FIG. *D*1

The output current I in the general two-mesh ladder structure shown in Fig. *D*1 is given (18) by

$$I_1 = \frac{EZ_2}{2Z_2\left(Z + \frac{Z_1}{2}\right) + \left(Z_1 + \frac{Z_1}{2}\right)^2} \tag{1}$$

$$= \frac{E}{2\left(Z + \frac{Z_1}{2}\right) + \frac{1}{Z_2}\left(Z + \frac{Z_1}{2}\right)^2}. \tag{2}$$

The input current I_0 is given (18) by

$$I_0 = \frac{E\left(Z_2 + Z + \frac{Z_1}{2}\right)}{2Z_2\left(Z + \frac{Z_1}{2}\right) + \left(Z + \frac{Z_1}{2}\right)^2}. \tag{3}$$

The parameters of a constant-k low-pass T wave-filter section with $R = k$ terminations are:

$$k = R = \sqrt{\frac{L_1}{C_2}}, \qquad \omega_c = \frac{2}{\sqrt{L_1 C_2}}, \qquad L_1 = \frac{2R}{\omega_c}, \qquad C_2 = \frac{2}{R\omega_c},$$

$$Z_1 = pL = \frac{2R}{\omega_c}p, \qquad Z_2 = \frac{1}{pC_2} = \frac{R\omega_c}{2p}, \qquad Z = k = R.$$

Substituting these values in Eq. (3),

$$I_0 = \frac{E\left(R\frac{\omega_c}{2p} + R + R\frac{p}{\omega_c}\right)}{\frac{R\omega_c}{p}\left(R + R\frac{p}{\omega_c}\right) + \left(R + R\frac{p}{\omega_c}\right)^2}$$

$$= \frac{\omega_c E}{2R}\,\frac{2p^2 + 2\omega_c p + \omega_c^2}{(p^2 + \omega_c p + \omega_c^2)(p + \omega_c)}. \tag{4}$$

Letting

$$K \equiv \frac{\omega_c}{2R},$$

it follows that

$$Y_0(p) = \frac{1}{Z_0(p)} = K\,\frac{2p^2 + 2\omega_c p + \omega_c^2}{(p + \omega_c)(p^2 + \omega_c p + \omega_c^2)}. \tag{5}$$

Applying the Heaviside Expansion Theorem (35),

$$I_0(t) = \left(\frac{1}{Z_0(0)} + \sum_{p=p_1}^{p=p_n} \frac{\epsilon^{pt}}{pZ_0'(p)}\right) E\mathbf{1}. \tag{6}$$

$$Z_0(p) = 0, \qquad (p + \omega_c)(p^2 + \omega_c p + \omega_c^2) = 0,$$

$$p_1 = -\omega_c, \qquad p_2 = \frac{-\omega_c + \sqrt{\omega_c^2 - 4\omega_c^2}}{2} = \frac{\omega_c}{2}(-1 + j\sqrt{3}),$$

$$p_3 = \frac{\omega_c}{2}(-1 - j\sqrt{3}), \qquad Z_0(0) = \frac{K}{\omega_c} = \frac{1}{2R}.$$

$$\frac{1}{pZ_0'(p)} = K\,\frac{2p^2 + 2\omega_c p + \omega_c^2}{p(3p^2 + 4\omega_c p + 2\omega_c^2)}.$$

Substituting these values in Eq. (6), collecting terms, and substituting the value of K,

$$I_0(t) = \frac{E}{2R}\left[1 - \epsilon^{-\omega_c t} + \frac{2\epsilon^{-(\omega_c/2)t}}{\sqrt{3}}\left(\frac{\epsilon^{j(\sqrt{3}/2)t} - \epsilon^{-j(\sqrt{3}/2)t}}{2j}\right)\right]\mathbf{1}.$$

So, by a well-known trigonometric identity,

$$I_0(t) = \frac{E}{2R}\left(1 - \epsilon^{-\omega_c t} + \frac{2\sqrt{3}}{3}\epsilon^{-(\omega_c/2)t}\sin\frac{\sqrt{3}}{2}t\right)\mathbf{1}. \qquad (7)$$

Eq. (7) gives the input current to the section when an $E\mathbf{1}$ wave of electromotive force is impressed. The driving-point time admittance is, of course,

$$A_0(t) = \frac{1}{2R}\left(1 - \epsilon^{-\omega_c t} + \frac{2\sqrt{3}}{3}\epsilon^{-(\omega_c/2)t}\sin\frac{\sqrt{3}}{2}t\right)\mathbf{1}. \qquad (8)$$

APPENDIX *E*

A Comparison of the Driving-Point Time Admittances of Low- and High-Pass Series *M*-Type Wave-Filter Sections with Those of the Corresponding Prototype Sections

In Fig. *E*1 are given the parameters of the driving-point time admittances of low- and high-pass prototype and series *M*-type

Low-Pass

	Prototype	Series *M*-type
Radial Frequency.....	$\omega_p = \frac{\sqrt{3}\omega_c}{2}$,	$\omega_m = \sqrt{\frac{4c + 3m}{4m(c + m^2)}}\,\omega_c$
Amplitude Factor.....	$\alpha_p = \frac{\sqrt{3}}{3R}$,	$\alpha_m = \sqrt{\frac{m^2}{1 + 2m^2}}\,\frac{1}{R}$
Decrement Coefficient...	$b_p = \frac{\omega_c}{2}$,	$b_m = \frac{\omega_c}{2(c + m)}$.

High-Pass

	Prototype	Series *M*-type
Radial Frequency.....	$\omega_p = \frac{\sqrt{3}\omega_c}{2}$,	$\omega_m = \sqrt{\frac{4 - m^2}{4}}\,\omega_c$
Amplitude Factor.....	$\alpha_p = \frac{\sqrt{3}}{3R}$,	$\alpha_m = \frac{m}{\sqrt{4 - m^2}R}$
Decrement Coefficient...	$b_p = \frac{\omega_c}{2}$,	$b_m = \frac{m\omega_c}{2}$.

FIG. *E*1

Variation of amplitude-factor, decrement coefficient, and radial frequency with m, of the driving-point time admittance of a low-pass series M-type wave-filter section.

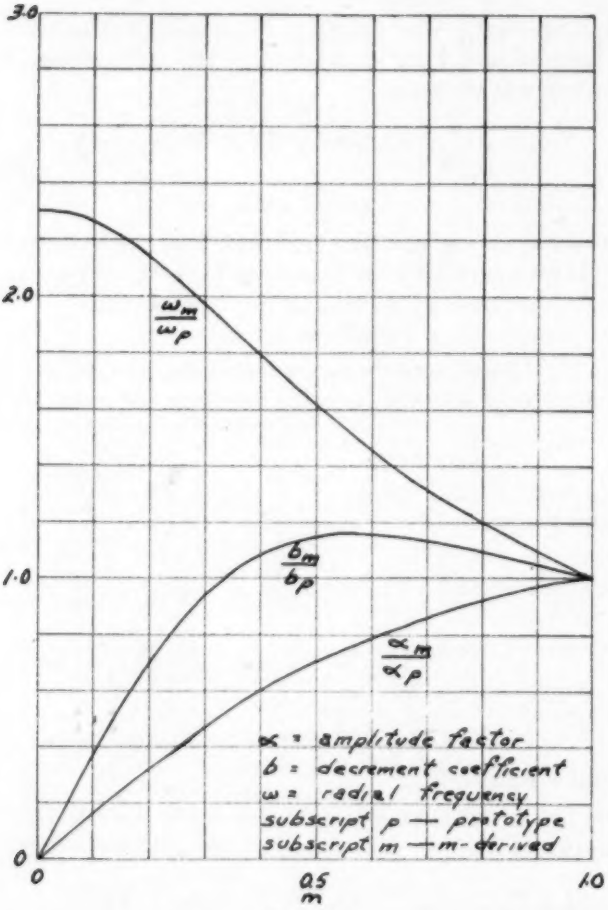


FIG. E2

Variation of amplitude-factor, decrement coefficient, and radial frequency with m, of the driving-point time admittance of a high-pass series M-type wave-filter section.

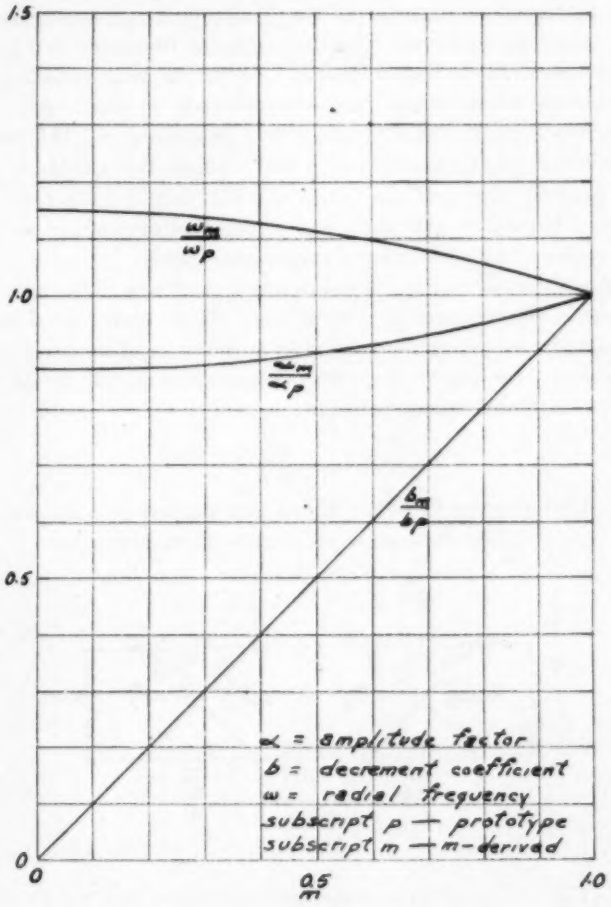


FIG. E3

wave-filter sections. The variations of the parameters for the M-type sections with respect to m are shown in Figs. $E2$ and $E3$.

An inspection of Fig. $E2$ shows that as m varies from 1 to 0, the frequency of oscillation of the driving-point time admittance varies almost uniformly from the prototype frequency to a little more than double that frequency. With the same variation of m the amplitude factor decreases uniformly to zero, while the decrement coefficient rises slowly to a maximum of 1.155 times the prototype coefficient at $m = 0.577$, passes through the value of the prototype coefficient at $m = 0.334$, then drops rapidly to zero. It may be seen that over a considerable range of m the decrement coefficient is not changed appreciably.

An inspection of Fig. $E3$ shows effects markedly different from those in the low-pass M-type section. Here, as m varies from 1 to 0, the frequency rises slowly from the value of the prototype frequency, the amplitude factor decreases slowly, and the decrement coefficient decreases linearly to zero.

APPENDIX F

Solution for the Current in the nth Section of a Balanced Constant-Resistance Lattice Type Network

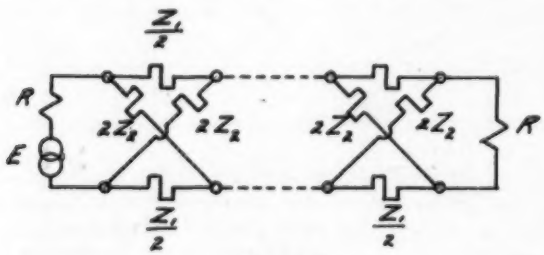


Fig. $F1$

In Fig. $F1$, $\frac{Z_1}{2}$ and $2Z_2$ are the arm impedances, and $\sqrt{Z_1 Z_2} = R$ is the iterative impedance.

Let P denote the propagation constant per section of the structure, and I_n the current crossing the junction point between the nth and the $(n+1)$st sections. Then I_0 will denote the current entering the structure.

Now it is shown in (18) that

$$I_n = I_0 \epsilon^{-nP}, \tag{1}$$

and

$$I_0 = \frac{E}{2\sqrt{Z_1 Z_2}}. \tag{2}$$

Also, from (4) it is known that

$$nP = ln\left(\frac{1+\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}{1-\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}\right)^n. \tag{3}$$

The relation in Eq. (3) may be written in the form

$$nP = ln\left(\frac{1-\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}{1+\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}\right)^{-n}. \tag{4}$$

So

$$\epsilon^{-nP} = \left(\frac{1\mp\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}{1\pm\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}\right)^{\pm n}. \tag{5}$$

Substituting values from Eqs. (5) and (2) in (1),

$$I_n = \frac{E}{2\sqrt{Z_1 Z_2}}\left(\frac{1\mp\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}{1\pm\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}\right)^{\pm n}. \tag{6}$$

So

$$Y_n(p) = \frac{1}{2\sqrt{Z_1 Z_2}}\left(\frac{1\mp\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}{1\pm\frac{1}{2}\sqrt{\frac{Z_1}{Z_2}}}\right)^{\pm n}. \tag{7}$$

Bibliography

1. "A General Expansion Theorem for the Transient Oscillations of a Connected System," by J. R. Carson, *Phys. Rev.*, vol. 10, pp. 217–225, Sept. 1917.
2. "Theory of the Transient Oscillations of Electrical Networks and Transmission Systems," by J. R. Carson, *Proc. A. I. E. E.*, vol. 28, No. 3, pp. 407–489, March 1919.
3. "Electric Wave Filters," by F. B. Jewett, *Q. S. T.*, Oct. 1922.
4. "Physical Theory of the Electric Wave-Filter," by G. A. Campbell, *B. S. T. Jl.*, vol. I, No. 2, pp. 1–32, Nov. 1922.
5. "Principles of Alternating Currents," by R. R. Lawrence, McGraw-Hill Book Co., 1922.
6. "Theory and Design of Uniform and Composite Electric Wave-Filters," by O. J. Zobel, *B. S. T. Jl.*, vol. II, No. 1, pp. 1–46, Jan. 1923.
7. "Transient Oscillations in Electric Wave-Filters," by J. R. Carson and O. J. Zobel, *B. S. T. Jl.*, vol. II, No. 3, pp. 1–52, July 1923.
8. "Electric Filters," by F. S. Dellenbaugh, Jr., *Q. S. T.*: Part I, July 1923, p. 15; Part II, Aug. 1923, p. 18.
9. "Die Ausbildung von dauernden Sinusschwingungen in einem langen homogenen Kabel," by F. Lüschen and K. Küpfmüller, *W. V. S. K.*, Bd. III, March 1923.
10. "A Reactance Theorem," by R. M. Foster, *B. S. T. Jl.*, vol. III, No. 2, pp. 259–267, April 1924.
11. "Transmission Characteristics of Electric Wave-Filters," by O. J. Zobel, *B. S. T. Jl.*, vol. III, No. 4, pp. 567–620, Oct. 1924.
12. "Theorems Regarding the Driving-Point Impedance of Two-Mesh Circuits," by R. M. Foster, *B. S. T. Jl.*, vol. III, No. 4, pp. 651–685, Oct. 1924.
13. "Über Einschwingvorgänge in Wellenfiltern," by K. Küpfmüller, *E. N. T.*, Bd. 1, Heft 5, 1924.
14. "A Note on the Theory of Artificial Telephone and Transmission Lines," by A. C. Bartlett, *Phil. Mag.*, vol. 48, pp. 859–865, 1924.
15. "Mutual Inductance in Wave Filters with an Introduction on Filter Design," by K. S. Johnson and T. E. Shea, *B. S. T. Jl.*, vol. IV, No. 1, pp. 52–111, Jan. 1925.
16. "Selective Circuits and Static Interference," by J. R. Carson, *B. S. T. Jl.*, vol. IV, No. 2, pp. 265–279, April 1925.
17. "A Class of Equivalent Electrical Networks," by A. C. Bartlett, *Phil. Mag.*, vol. 49, pp. 728–739, 1925.
18. "Transmission Circuits for Telephonic Communication," by K. S. Johnson, D. Van Nostrand Co., 1st Ed., 1925.
19. "Electric Circuit Theory and the Operational Calculus," by J. R. Carson, McGraw-Hill Book Co., 1st Ed., 1926: and *B. S. T. Jl.*, vol. IV, No. 4, pp. 685–761, Oct. 1925; vol. V, No. 1, pp. 50–95, Jan. 1926; vol. V, No. 2, pp. 336–384, April 1926.

20. "Über Einschwingvorgänge in Pupinleitungen und ihre Verminderung," by K. Küpfmüller and H. F. Mayer, *W. V. S. K.*, Bd. V, March 1926.
21. "Die Erhöhung der Reichweite von Pupinleitung durch Echoserrung und Phasenausgleich," by K. Küpfmüller, *E. N. T.*, Bd. 3, Heft 3, 1926.
22. "Störungsverminderung durch selektive Schaltmittel beim drahtlosen Empfang," by K. Küpfmüller, *E. N. T.*, Bd. 3 Heft 3, 1926.
23. "Properties of the Generalized Artificial Line," by A. C. Bartlett, *Phil. Mag.*, vol. 17, pp. 553–557, 1926.
24. "Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit," by W. Cauer, *Archiv. f. Elektrot.*, Bd. 17, Heft 4, S. 355–388, Dec. 1926.
25. "Über die Wahl der Trägerfrequenzen für die Tonfrequenztelegraphie," by F. Lüschen and K. Küpfmüller, *E. N. T.*, Bd. 4, Heft 4, April 1927.
26. "Eine Rechentafel zur Berechnung von Schwingungskreisen," by F. Strecker, *W. V. S. K.*, Bd. VII, Sept. 1927.
27. "Über die Variablen eines Passiven Vierpols," by W. Cauer, *Sitz. d. Preuss. Ak. d. Wiss.*, Bd. 6, pp. 268–274, Dec. 1927.
28. "Heaviside's Operational Calculus as Applied to Electricity and Physics," by E. J. Berg, *G. E. Review*, Dec. 1927; Feb. to May, July to Sept. 1928.
29. "Über den Einfluss des Phasenmasses und der Dämfung bei der Übertragung von modulierten Wellen," by H. Bartels, *W. V. S. K.*, Bd. VII, Jan. 1928.
30. "Über Beziehungen Zwischen Frequenzcharakteristiken und Ausgleischvorgängen in linearen Systemen," by K. Küpfmüller, *E. N. T.*, Bd. 5, Heft 1, 1928.
31. "Phase Distortion and Phase Distortion Correction," by S. P. Mead, *B. S. T. Jl.*, vol. VII, No. 2, pp. 195–224, April 1928.
32. "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," by O. J. Zobel, *B. S. T. Jl.*, vol. VII, No. 3, pp. 438–534, July 1928.
33. "The Practical Application of the Fourier Integral," by G. A. Campbell, *B. S. T. Jl.*, vol. VII, No. 4, pp. 639–707, Oct. 1928.
34. "A New Method for Obtaining Transient Solutions of Electrical Networks," by W. P. Mason, *B. S. T. Jl.*, vol. VIII, No. 1, pp. 109–134, Jan. 1929.
35. "Operational Circuit Analysis," by V. Bush, John Wiley & Sons, 1929.
36. "Vector Presentation of Broad-Band Wave Filters," by R. F. Mallina and O. Knackmuss, *A. I. E. E. Jl.*, vol. 47, No. 4, pp. 264–269, April 1929.
37. "The Use of Continued Fractions in the Design of Electric Networks," by T. C. Fry, *Bul. of Am. Math. Soc.*, vol. 35, pp. 463–498, 1929.
38. "Transmission Networks and Wave Filters," by T. E. Shea, D. Van Nostrand Co., 1929.
39. "Heaviside's Operational Calculus as Applied to Engineering and Physics," by E. J. Berg, McGraw-Hill Book Co., 1929.
40. "Notes on the Heaviside Operational Calculus," by J. R. Carson, *B. S. T. Jl.*, vol. IX, No. 1, pp. 150–162, Jan. 1930.

41. "A Method of Impedance Correction," by H. W. Bode, *B. S. T. Jl.*, vol. IX, No. 4, pp. 794–835, Oct. 1930.
42. "Impedance Correction of Wave-Filters," by E. B. Payne, *B. S. T. Jl.*, vol. IX, No. 4, pp. 770–793, Oct. 1930.
43. "Die Siebschaltungen der Fernmeldtechnik," by W. Cauer, *Zeits. f. Angewandte Math. und Mech.*, Bd. 10, Heft 5, S. 425–433, Oct. 1930.
44. "Extensions to the Theory and Design of Electric Wave-Filters," by O. J. Zobel, *B. S. T. Jl.*, vol. X, No. 2, pp. 284–341, April 1931.
45. "Group Theory and the Electric Circuit," by N. Howitt, *Phys. Rev.*, vol. 37, No. 12, pp. 1583–1595, June 15, 1931.
46. "The Statistical Energy-Frequency Spectrum of Random Disturbances," by J. R. Carson, *B. S. T. Jl.*, vol. X, No. 3, pp. 374–381, July 1931.
47. "Untersuchungen über ein Problem, das drei positiv definite quadratische Formen mit Streckenkomplexe in Beziehung setzt," by W. Cauer, *Math. Annalen*, Bd. 105, Heft 1, S. 86–132, July 1931.
48. "Synthesis of a Finite Two-terminal Network whose Driving-point Impedance is a Prescribed Function of Frequency," by O. Brune, *M. I. T. Jl. of Math. and Phys.*, vol. X, No. 3, pp. 191–236, Aug. 1931.
49. "Fourier Integrals for Practical Applications," by G. A. Campbell and R. M. Foster, Bell Telephone System Monograph B-584, Sept. 1931.
50. "Equivalent Electrical Networks," (Letter to Editor) by O. Brune, *Phys. Rev.*, vol. 38, No. 9, p. 1783, Nov. 1, 1931.

THE INVERSE MATRIX FOR DE LA VALLÉE-POUSSIN SUMMATION [1]

By George Rutledge

1. **Introduction.** We begin by stating certain known definitions and facts concerning Cesàro and de la Vallée-Poussin summation.

Given the series

$$u_0 + u_1 + u_2 + u_3 + \cdots, \tag{1}$$

and the expressions

$$C_n^{(k)} = u_0 + \sum_{j=1}^{n} \frac{\binom{n-j+k}{k}}{\binom{n+k}{k}} u_j, \tag{2}$$

$$V_n = u_0 + \sum_{j=1}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} u_j, \tag{3}$$

we say that the series (1) is summable Cesàro (C_k) if $\lim_{n\to\infty} C_n^{(k)}$ exists, and that it is summable de la Vallée-Poussin if $\lim_{n\to\infty} V_n$ exists. Existence of a limit for (2) implies existence of the same limit for (3), but this statement does not hold conversely.[2]

The importance of the particularly simple form of the solution of equations (2) for the quantities u_j has been noted by Hardy.[3] It will serve our purpose to exhibit this solution [4] by means of

[1] Presented to the American Mathematical Society, March 25, 1932.

[2] Gronwall, *Journal für die reine und angewandte Mathematik*, 147 (1917), p. 16.

[3] *Proceedings of the London Mathematical Society*, (2), 6 (1907), p. 257.

[4] Hardy, loc. cit., p. 258.

the matrix of coefficients of the equations (2), which we shall denote by (C_k), and the inverse matrix $(C_k)^{-1}$ as follows:

MATRIX (C_k)

$$\begin{Vmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & \dfrac{\binom{k}{k}}{\binom{1+k}{k}} & 0 & 0 & \cdots \\ 1 & \dfrac{\binom{1+k}{k}}{\binom{2+k}{k}} & \dfrac{\binom{k}{k}}{\binom{2+k}{k}} & 0 & \cdots \\ 1 & \dfrac{\binom{2+k}{k}}{\binom{3+k}{k}} & \dfrac{\binom{1+k}{k}}{\binom{3+k}{k}} & \dfrac{\binom{k}{k}}{\binom{3+k}{k}} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{Vmatrix}. \tag{4}$$

MATRIX $(C_k)^{-1}$

$$\begin{Vmatrix} \binom{1+k}{0}\binom{k}{k} & 0 & 0 & 0 & \cdots \\ -\binom{1+k}{1}\binom{k}{k} & \binom{1+k}{0}\binom{1+k}{k} & 0 & 0 & \cdots \\ \binom{1+k}{2}\binom{k}{k} & -\binom{1+k}{1}\binom{1+k}{k} & \binom{1+k}{0}\binom{2+k}{k} & 0 & \cdots \\ -\binom{1+k}{3}\binom{k}{k} & \binom{1+k}{2}\binom{1+k}{k} & -\binom{1+k}{1}\binom{2+k}{k} & \binom{1+k}{0}\binom{3+k}{k} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{Vmatrix}. \tag{5}$$

That the product of the matrices (4) and (5) is the unit matrix is readily verified by means of the identity

$$\sum_{s=0} (-1)^s \binom{m+k-s}{k}\binom{1+k}{s} = 0, \quad m > 0. \tag{6}$$

The expression (6) is the coefficient of x^m in $(1-x)^{-1-k}(1-x)^{1+k}$. We note that each column of $(C_k)^{-1}$ contains only $2+k$ non-zero elements.

It is the purpose of this paper to show that the solution of equations (3) is likewise simple and elegant in form. The equations (3) and the inverse equations will be exhibited by means of matrices analogous to (4) and (5), to be denoted by (VP) and $(VP)^{-1}$, respectively.

As an application, the de la Vallée-Poussin sequence for the series $\sum_{j=1}^{\infty} v_j u_j$ will be exhibited as a linear transformation of the de la Vallée-Poussin sequence for the series $\sum_{j=1}^{\infty} u_j$.

2. **Inverse of the (VP) Matrix.** The matrix of coefficients of the equations (3) is

MATRIX (VP)

$$\left\| \begin{matrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & \dfrac{\binom{2}{0}}{\binom{2}{1}} & 0 & 0 & \cdots \\ 1 & \dfrac{\binom{4}{1}}{\binom{4}{2}} & \dfrac{\binom{4}{0}}{\binom{4}{2}} & 0 & \cdots \\ 1 & \dfrac{\binom{6}{2}}{\binom{6}{3}} & \dfrac{\binom{6}{1}}{\binom{6}{3}} & \dfrac{\binom{6}{0}}{\binom{6}{3}} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix} \right\| . \tag{7}$$

The de la Vallée-Poussin matrix involves no parameter analogous to k of the Cesàro matrix.

We propose to prove that the inverse matrix is

MATRIX $(VP)^{-1}$

$$\left\| \begin{matrix} 1 & 0 & 0 & 0 & \cdots \\ -2 & 2\binom{1}{1}\binom{1}{1} & 0 & 0 & \cdots \\ 2 & -2\binom{2}{1}\binom{2}{1} & 2\binom{2}{2}\binom{3}{2} & 0 & \cdots \\ -2 & 2\binom{3}{1}\binom{3}{1} & -2\binom{3}{2}\binom{4}{2} & 2\binom{3}{3}\binom{5}{3} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix} \right\|. \quad (8)$$

That the elements of the product matrix $(VP)(VP)^{-1}$ below the principal diagonal in the first column are zero is an immediate consequence of the identity $(1-1)^{2n}=0$.

It then remains to be proved that the matrices obtained by deleting the first row and the first column of each of the matrices (7) and (8) are inverse matrices. These are the matrices corresponding to summation of the series

$$u_1+u_2+u_3+\cdots, \quad (9)$$

by means of the limit of the expression

$$V_n'=\sum_{j=1}^{n}\frac{\binom{2n}{n-j}}{\binom{2n}{n}}u_j. \quad (10)$$

Since the expression (10) and its limit arise *naturally*[5] in problems associated with the Stirling interpolation series, and since in what is to follow we are concerned with the intimate relation between de la Vallée-Poussin summation and Stirling series expansions, we shall find it convenient to use the notation (VP'), $(VP')^{-1}$ for the matrices (7) and (8), respectively, with the first row and the first column of each deleted; also to refer

[5] Rutledge, this JOURNAL, v. 9 (1930), p. 261.

to the limit of V_n' as the (VP')-sum of the series (9). Clearly, if the series (9) has (VP')-sum s, then the series (1) has (VP)-sum $u_0 + s$, and conversely.

To show that the matrices (VP') and $(VP')^{-1}$ are inverse matrices we resort to the following considerations.

If $F(x)$ is an even integral function which takes the value zero for $x = 0$, and if

$$\sum_{j=1}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} (-1)^{j-1} \frac{2F(j)}{j^2} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \cdots + \frac{1}{\lambda_n}, \quad (11)$$

where $\frac{1}{\lambda_k}$, $k = 1, 2, \cdots$, is independent of n, then the formal expansion of $F(x)$ in Stirling's series is[6]

$$\frac{1}{\lambda_1} x^2 + \frac{1}{\lambda_2} x^2 \left(1 - \frac{x^2}{1^2}\right) + \frac{1}{\lambda_3} x^2 \left(1 - \frac{x^2}{1^2}\right)\left(1 - \frac{x^2}{2^2}\right) + \cdots. \quad (12)$$

The (VP')-sum of the series

$$\frac{2F(1)}{1^2} - \frac{2F(2)}{2^2} + \frac{2F(3)}{3^2} - \cdots \quad (13)$$

is the sum in the ordinary sense of

$$\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \frac{1}{\lambda_3} + \cdots, \quad (14)$$

provided the latter is convergent.

The following rational integral functions are expressed by means of Stirling series reducing to two terms each.

$$x^2 \binom{x}{1}\binom{x}{1} = 1^2 \binom{x}{1}\binom{x}{1} + 2^2 \binom{x+1}{2}\binom{x}{2}$$
$$= x^2 - x^2 \left(1 - \frac{x^2}{1^2}\right),$$

[6] Rutledge, loc. cit., p. 262; also see R. D. Douglass, this JOURNAL, v. 10 (1931), p. 131.

$$x^2\binom{x+1}{2}\binom{x}{2} = 2^2\binom{x+1}{2}\binom{x}{2} + 3^2\binom{x+2}{3}\binom{x}{3}$$
$$= -x^2\left(1-\frac{x^2}{1^2}\right) + x^2\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right), \quad (15)$$
$$x^2\binom{x+2}{3}\binom{x}{3} = 3^2\binom{x+2}{3}\binom{x}{3} + 4^2\binom{x+3}{4}\binom{x}{4}$$
$$= x^2\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right)$$
$$- x^2\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right)\left(1-\frac{x^2}{3^2}\right),$$
. . . . . .

The corresponding λ-series (14) are, respectively,

$$\begin{aligned} &1 - 1 + 0 + 0 + 0 + \cdots, \\ &0 - 1 + 1 + 0 + 0 + \cdots, \\ &0 + 0 + 1 - 1 + 0 + \cdots, \\ &\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \end{aligned} \quad (16)$$

Then by application of (11) we have the following identities, which are precisely the identities needed to prove that all elements below the principal diagonal in the product matrix $(VP')(VP')^{-1}$ are zero.

$$\sum_{j=1}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} (-1)^{j-1} 2\binom{j}{1}\binom{j}{1} = 0, \quad n > 1,$$
$$\sum_{j=2}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} (-1)^{j-1} 2\binom{j}{2}\binom{j+1}{2} = 0, \quad n > 2, \quad (17)$$
$$\sum_{j=3}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} (-1)^{j-1} 2\binom{j}{3}\binom{j+2}{3} = 0, \quad n > 3,$$
. . . . . .

Moreover, we have also from (11)

$$\frac{\binom{2}{0}}{\binom{2}{1}} 2\binom{1}{1}\binom{1}{1} = 1, \qquad \frac{\binom{4}{0}}{\binom{4}{2}} 2\binom{2}{2}\binom{3}{2} = 1, \qquad \cdots. \tag{18}$$

This completes the proof that (VP') and $(VP')^{-1}$ are inverse matrices and also that the complete matrices (7) and (8) are inverse matrices.

3. **Transformation of the (VP')-sequence.** If

$$s_n' = \sum_{j=1}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} u_j, \tag{19}$$

and if

$$s_n'' = \sum_{j=1}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} \nu_j u_j, \tag{20}$$

then it is clear that the variables s_n' are transformed into the variables s_n'' by a linear transformation which is independent of the u's. The matrix of this linear transformation is

$$(VP')(\nu)(VP')^{-1}, \tag{21}$$

where

$$(\nu) = \begin{Vmatrix} \nu_1 & 0 & 0 & \cdots \\ 0 & \nu_2 & 0 & \cdots \\ 0 & 0 & \nu_3 & \cdots \\ \cdot & \cdot & \cdot & \cdot \end{Vmatrix}. \tag{22}$$

Then

$$s_n'' = c_{n1}s_1' + c_{n2}s_2' + \cdots + c_{nn}s_n', \tag{23}$$

where

$$c_{n1} = \frac{\binom{2n}{n-1}}{\binom{2n}{n}} 2\binom{1}{1}\binom{1}{1}\nu_1$$

$$-\frac{\binom{2n}{n-2}}{\binom{2n}{n}}2\binom{2}{1}\binom{2}{1}\nu_2+\frac{\binom{2n}{n-3}}{\binom{2n}{n}}2\binom{3}{1}\binom{3}{1}\nu_3-\cdots$$

$$c_{n2}=\frac{\binom{2n}{n-2}}{\binom{2n}{n}}2\binom{2}{2}\binom{3}{2}\nu_2-\frac{\binom{2n}{n-3}}{\binom{2n}{n}}2\binom{3}{2}\binom{4}{2}\nu_3+\cdots \quad (24)$$

$$c_{n3}=\frac{\binom{2n}{n-3}}{\binom{2n}{n}}2\binom{3}{3}\binom{5}{3}\nu_3-\cdots$$

. . . . . .

If in (6) we place $m = 1 + k = j$, we obtain

$$\sum_{s=0}^{j-1}(-1)^s\binom{j}{j-s}\binom{2j-1-s}{j-s}=(-1)^{j-1}. \quad (25)$$

From this it follows that

$$c_{n1}+c_{n2}+c_{n3}+\cdots+c_{nn}=\sum_{j=1}^{n}\frac{\binom{2n}{n-j}}{\binom{2n}{n}}(-1)^{j-1}2\nu_j. \quad (26)$$

The necessary and sufficient conditions that s_n'', as defined by (23), shall have a finite limit in all cases in which s_n' has a finite limit have been stated by Kojima.[7] These conditions are three in number, as follows:

(1) $\lim_{n\to\infty} c_{nk}$ exists for each k;

(2) $\lim_{n\to\infty}(c_{n1}+c_{n2}+\cdots+c_{nn})$ exists;

(3) $|c_{n1}|+|c_{n2}|+|c_{n3}|+\cdots+|c_{nn}|<M$,

where M is a constant independent of n.

[7] *Tôhoku Mathematical Journal*, v. 12 (1917), pp. 291–326.

The second of these conditions applied to (26) leads to the following statement, which is also directly evident.

If $\nu_1, \nu_2, \nu_3, \cdots$ are constants such that the series $\sum \nu_i u_i$ is (VP')-summable whenever the series $\sum u_i$ is (VP')-summable, then the series $\sum (-1)^{i-1}\nu_i$ is (VP')-summable.

By reference to (15) we have

$$\begin{aligned}
&\binom{x}{1}\binom{x}{1} = \frac{x^2}{1^2},\\
&\binom{x+1}{2}\binom{x}{2} = -\frac{x^2}{2^2}\left(1-\frac{x^2}{1^2}\right), \qquad (27)\\
&\binom{x+2}{3}\binom{x}{3} = \frac{x^2}{3^2}\left(1-\frac{x^2}{1^2}\right)\left(1-\frac{x^2}{2^2}\right),\\
&\cdots\cdots\cdots\cdots
\end{aligned}$$

The corresponding λ-series (14) are, respectively,

$$\begin{aligned}
&\frac{1}{1^2}+0+0+0+0+\cdots,\\
&0-\frac{1}{2^2}+0+0+0+\cdots, \qquad (28)\\
&0+0+\frac{1}{3^2}+0+0+\cdots,\\
&\cdots\cdots\cdots\cdots
\end{aligned}$$

Then by application of (11) we have

$$\begin{aligned}
&\sum_{j=1}^{n}\frac{\binom{2n}{n-j}}{\binom{2n}{n}}(-1)^{j-1}2\binom{j}{1}\binom{j}{1}\frac{1}{j^2} = \frac{1}{1^2},\\
&\sum_{j=2}^{n}\frac{\binom{2n}{n-j}}{\binom{2n}{n}}(-1)^{j-2}2\binom{j}{2}\binom{j+1}{2}\frac{1}{j^2} = \frac{1}{2^2}, \qquad (29)
\end{aligned}$$

$$\sum_{j=3}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} (-1)^{j-3} 2 \binom{j}{3} \binom{j+2}{3} \frac{1}{j^2} = \frac{1}{3^2},$$

. . . . . .

The sums (29) are precisely the sums (24) with $\nu_j = 1/j^2$. Hence if s_n', s_n'' denote, respectively, the (VP')-sequences for the series $\sum u_j$ and $\sum (1/j^2)u_j$, then (23) becomes [8]

$$s_n'' = \frac{1}{1^2} s_1' + \frac{1}{2^2} s_2' + \cdots + \frac{1}{n^2} s_n'. \tag{30}$$

Since the coefficients in (30) satisfy the conditions of Kojima, it follows that whenever the series $\sum u_j$ is (VP')-summable the series $\sum (1/j^2)u_j$ is likewise (VP')-summable.

Moreover it is by virtue of this remarkably simple equation (30) that we have the formula [9]

$$\lim_{n \to \infty} \sum_{j=1}^{n} \frac{\binom{2n}{n-j}}{\binom{2n}{n}} u_j = \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{s=1}^{k} \frac{\binom{2k}{k-s}}{\binom{2k}{k}} s^2 u_s, \tag{31}$$

whereby a series $\sum u_j$ which is (VP')-summable to a certain value is transformed into a series $\sum 1/\lambda_k$ which is convergent in the ordinary sense to this same value, and whereby a convergent series $\sum u_j$ is transformed into an equivalent convergent series $\sum 1/\lambda_k$.

[8] The equation (30) has been previously obtained by the writer by a less general method. Rutledge, loc. cit., p. 270.

[9] R. D. Douglass, loc. cit., p. 134, (7), (8).

SYNTHESIS OF ELECTRIC NETWORKS BY MEANS OF THE FOURIER TRANSFORMS OF LAGUERRE'S FUNCTIONS

By Y. W. Lee

This paper presents a new method in electric network synthesis. Continued fractions have played a role in earlier attacks on this problem.[1] Here will be presented a new method of attack based on the use of the Fourier transforms [2] of Laguerre's functions.

While to represent the behavior of a network by a formal expression and to give it a physical interpretation is now almost a matter of routine, the reverse process, that is, the finding of a physically realizable network whose properties correspond to a specified explicit formal expression, may not be accomplished with the same ease. When the preassigned properties are represented by graphs the difficulty is even greater.

We shall not go into the details of the present methods in the design of electric networks whose frequency characteristics are

[1] T. C. Fry, "The Use of Continued Fract ionsin the Design of Electrical Networks," *Bull. American Mathematical Society*, vol. XXXV, pp. 463–498, July–August, 1929; U. S. Patent 1,570,215, issued Jan. 19, 1926.

W. Cauer, "Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit," *Archiv fur Elektrotechnik*, vol. 17, 1926, pp. 355–388.

A. C. Bartlett, "A Note on the Theory of Artificial Telephone and Transmission Lines," *Philosophical Magazine*, vol. 48, 1924, p. 859; "Properties of the Generalized Artificial Line," *Phil. Mag.*, vol. 1, 1926, p. 553; British Patent 290,701, issued 1928.

[2] N. Wiener, "Generalized Harmonic Analysis," *Acta Mathematica*, vol. 55, 1930, pp. 117–258. This excellent paper contains a comprehensive and classified bibliography.

V. Bush, "Operational Circuit Analysis," John Wiley and Sons, 1929, pp. 165–188, and pp. 366–379.

G. A. Campbell, "The Practical Application of the Fourier Integral," *Bell System Technical Journal*, vol. VII, 1928, pp. 639–707.

preassigned, but we can get a general idea of the matter from the following quotation from a recent paper by O. J. Zobel:[3]

"It would be most gratifying to be able to obtain directly from a desired propagation characteristic the corresponding form of network. This is generally a difficult problem and it becomes necessary to resort to simplifying methods somewhat similar to those employed in the design of electric wave-filters. . . . We would, therefore, begin with known forms of networks whose general propagation characteristics have been determined and choose from them one or more whose combination offers the possibility of giving a satisfactory desired result."

The properties of a network may be expressed in various ways. Here we shall confine ourselves primarily to the use of admittances either as transfer values or driving-point values as the case may be. The specification of the properties of a network may be made in terms of any one of the four parameters of the admittance; namely, the conductance, the susceptance, the modulus, and the phase. These parameters are functions of the frequency. As certain relations exist among these parameters, it is not generally possible to specify independently more than one property of a network over the entire frequency range $(0, \infty)$. The interrelations of the parameters of the admittance function of a network are not generally known. These interrelations, which form an important part in the present method of synthesis, shall be discussed in the first part of this paper.

It is sometimes desirable to synthesize a network with two or more specified properties. For instance, a network possessing a designable admittance both in modulus and in phase would find many applications which are increasing in importance as the problem of the improvement of quality in electrical transmission draws the attention of the electrical engineer more and more. However, the present investigation has not been extended to cover the multiple assignment of parameters.

[3] O. J. Zobel, "Distortion Correction in Electrical Circuits with Constant Resistance Recurrent Networks," *Bell System Technical Journal*, vol. VII, 1928, p. 466.

This paper covers the results of a doctoral dissertation which the author presented to the Massachusetts Institute of Technology in May, 1930. This research was suggested to the author by Dr. Vannevar Bush whose intense interest, many suggestions, and general guidance throughout the development constituted a most essential part in the undertaking. Dr. Norbert Wiener was the originator of the basic idea of this work; and he helped unreservedly in its mathematical theory. In particular, he suggested the use of the Fourier transforms of Laguerre's functions in the present method of synthesis, pointed out the application of Hilbert transforms to electric networks, and established the reciprocal relations between the modulus and the phase of an admittance function by a logarithmic transformation of the function. To these men the author offers his most appreciative thanks.

I

In the study of circuit transients, we have noticed the implicit relation between the conductance and the susceptance of a network. The explicit expression for the indicial admittance obtained by means of the Fourier integral is [4]

$$A(t) = P(0) + \frac{2}{\pi}\int_0^\infty \frac{Q(\omega)}{\omega}\cos \omega t d\omega$$

$$= \frac{2}{\pi}\int_0^\infty \frac{P(\omega)}{\omega}\sin \omega t d\omega, \qquad (1)$$

where

$A(t)$ = Indicial admittance.
$P(\omega)$ = Conductance of the network.
$Q(\omega)$ = Susceptance of the network.
t = Time.
ω = 2π times the frequency.
$P(0)$ = Conductance at zero frequency.

Differentiating with respect to t under the integral sign, we get

$$\int_0^\infty Q(\omega)\sin \omega t d\omega = -\int_0^\infty P(\omega)\cos \omega t d\omega. \qquad (2)$$

[4] V. Bush, loc. cit., p. 180.

This equation states that the Fourier cosine transform of the conductance is the negative sine transform of the susceptance of an admittance.

If we take the cosine transform of (2), we obtain

$$P(\omega) = -\frac{2}{\pi}\int_0^\infty \cos \omega u du \int_0^\infty Q(t) \sin utdt. \tag{3}$$

It is to be noted that P is an even function, and Q an odd function. Similarly, the sine transform of (2) is

$$Q(\omega) = -\frac{2}{\pi}\int_0^\infty \sin \omega u du \int_0^\infty P(t) \cos utdt. \tag{4}$$

Equation (3) states that the conductance is the negative cosine transform of the sine transform of the susceptance. Equation (4) states that the susceptance is the negative sine transform of the cosine transform of the conductance.

As far as the writer knows, the interrelations of the parameters of the admittance function have not been dealt with in electrical engineering literature. In mathematical literature however, we find a number of papers of recent years on "Conjugate Integrals," or "Hilbert Transforms." These transforms are immediately applicable in network problems. Going from (3) and (4) into the Hilbert transforms,[5] we find that

$$P(\omega) = \frac{1}{\pi}\int_{-\infty}^\infty \frac{Q(u)}{\omega - u} du \tag{5}$$

and

$$Q(\omega) = -\frac{1}{\pi}\int_{-\infty}^\infty \frac{P(u)}{\omega - u} du. \tag{6}$$

Or, written in another form

$$P(\omega) = \frac{1}{\pi}\int_0^\infty \frac{Q(\omega + u) - Q(\omega - u)}{u} du \tag{7}$$

[5] E. C. Titchmarsh, "Conjugate Trigonometrical Integrals," *Proceedings of the London Mathematical Society*, 2d series, vol. 24, 1926, pp. 109–130; "On Conjugate Functions," *Proc. London Math. Soc.*, 2d series, vol. 29, 1929, pp. 49–81.

F. M. Wood, "Reciprocal Integral Formulae," *Proc. London Math. Soc.*, 2d series, vol. 29, 1929, pp. 29–48.

and

$$Q(\omega) = -\frac{1}{\pi}\int_0^\infty \frac{P(\omega + u) - P(\omega - u)}{u}\,du. \tag{8}$$

These integrals are defined as principal values, that is, integrals of the type

$$\lim_{\epsilon \to 0} \int_\epsilon^\infty \cdot$$

The reciprocal relations (5) and (6) or (7) and (8) being known as Hilbert transforms or conjugate integrals, lead us to say that the conductance and the susceptance of the admittance of a network are Hilbert transforms or conjugate integrals of each other.

Since a non-dissipative circuit in parallel with a given circuit affects the susceptance alone, and a pure resistance in parallel with the same circuit amounts merely to the addition of a constant, the reciprocal relations of $P(\omega)$ and $Q(\omega)$ do not always hold. As inductances and capacitances are usually accompanied by resistances, the only exception to these relations in a physical network is therefore, when a circuit is shunted by a resistance. As is obvious, such a resistance can be easily accounted for in any actual problem. In regard to the convergence of the integrals, the reader is referred to the papers already cited.

An important relation is that between the modulus and the phase of an admittance function. To establish this, we shall consider the logarithm of the admittance. Since the admittance function $Y(\omega)$ is analytic over the right half of the complex plane, its logarithm enjoys the same property. Consequently, as

$$Y(\omega) = P(\omega) + jQ(\omega)$$

and

$$ln\ Y(\omega) = \tfrac{1}{2}ln(P^2 + Q^2) + j\tan^{-1}(Q/P),$$

where $j = \sqrt{-1}$, $\frac{1}{2}ln(P^2 + Q^2)$ bears to $\tan^{-1}(Q/P)$, or, the logarithm of the modulus bears to the phase, the relation which the conductance bears to the susceptance. In other words, the logarithm of the modulus of an admittance function and its phase

are Hilbert transforms of each other. We may write therefore, following (3) to (8), and letting

$$\phi = \tan^{-1}(Q/P), \qquad ln|Y| = \tfrac{1}{2}ln(P^2 + Q^2),$$

$$ln|Y| = -\frac{1}{\pi}\int_0^\infty \cos \omega u du \int_{-\infty}^\infty \phi(t) \sin utdt, \tag{9}$$

$$\phi = -\frac{1}{\pi}\int_0^\infty \sin \omega u du \int_{-\infty}^\infty ln|Y(t)| \cos utdt, \tag{10}$$

or

$$ln|Y| = \frac{1}{\pi}\int_{-\infty}^\infty \frac{\phi(u)}{\omega - u} du, \tag{11}$$

$$\phi = -\frac{1}{\pi}\int_{-\infty}^\infty \frac{ln|Y(u)|}{\omega - u} du, \tag{12}$$

or

$$ln|Y| = \frac{1}{\pi}\int_0^\infty \frac{\phi(\omega + u) - \phi(\omega - u)}{u} du, \tag{13}$$

$$\phi = -\frac{1}{\pi}\int_0^\infty \frac{ln|Y(\omega + u)| - ln|Y(\omega - u)|}{u} du. \tag{14}$$

Owing to the fact that a zero of a function is an infinity of its logarithm, the determination of phase by modulus or vice versa is not unique. It is this that permits the existence of phase-correcting networks of constant moduli over the frequency interval $(0, \infty)$.

Similar to the Parseval theorem for Fourier transforms we have the Parseval theorem for Hilbert transforms, which, when applied to networks, states that if two admittance functions are

$$Y_1(\omega) = P_1(\omega) + jQ_1(\omega)$$

and

$$Y_2(\omega) = P_2(\omega) + jQ_2(\omega),$$

then

$$\int_{-\infty}^\infty P_1(\omega)P_2(\omega)d\omega = \int_{-\infty}^\infty Q_1(\omega)Q_2(\omega)d\omega. \tag{15}$$

If $P_1 = P_2$ and $Q_1 = Q_2$, (15) becomes

$$\int_{-\infty}^{\infty} P^2(\omega)d\omega = \int_{-\infty}^{\infty} Q^2(\omega)d\omega. \tag{16}$$

This has been treated rigorously by Titchmarsh.[6]

In the following examples we shall apply the integrals which we have considered to a very simple circuit of resistance and inductance in series. To simplify it further, let the resistance be one ohm and the inductance one henry.

1. Given: $P(\omega) = \dfrac{1}{1+\omega^2}$. Find: $Q(\omega)$.

By equation (4) which is equivalent to (6)

$$\begin{aligned} Q(\omega) &= -\frac{2}{\pi}\int_0^{\infty} \frac{du}{(\omega-u)(1+u^2)} \\ &= -\frac{2}{\pi}\left[\frac{1}{1+\omega^2}\int_0^{\infty}\frac{du}{\omega-u} + \frac{1}{1+\omega^2}\int_0^{\infty}\frac{udu}{1+u^2} \right. \\ &\qquad \left. + \frac{\omega}{1+\omega^2}\int_0^{\infty}\frac{du}{1+u^2}\right] \\ &= -\frac{\omega}{1+\omega^2}. \end{aligned}$$

The result is recognized as the susceptance of the circuit under consideration.

2. Given: $Q(\omega) = -\dfrac{\omega}{1+\omega^2}$. Find: $P(\omega)$.

By equation (3) which is equivalent to (5)

$$\begin{aligned} P(\omega) &= \frac{2}{\pi}\int_0^{\infty}\frac{Q(u)}{\omega-u}du = -\frac{2}{\pi}\int_0^{\infty}\frac{udu}{(\omega-u)(1+u^2)} \\ &= -\frac{2}{\pi}\left[\frac{\omega}{1+\omega^2}\int_0^{\infty}\frac{du}{\omega-u} + \frac{\omega}{1+\omega^2}\int_0^{\infty}\frac{udu}{1+u^2} \right. \\ &\qquad \left. - \frac{1}{1+\omega^2}\int_0^{\infty}\frac{du}{1+u^2}\right] \\ &= \frac{1}{1+\omega^2}. \end{aligned}$$

[6] E. C. Titchmarsh, loc. cit., vol. 24, 1926, p. 119.

3. Given: $P(\omega) = \dfrac{1}{1+\omega^2}$. Find: $Q(\omega)$ by equation (4).

$$Q(\omega) = -\frac{2}{\pi}\int_0^\infty \sin \omega u du \int_0^\infty \frac{1}{1+t^2}\cos utdt \text{ [7]}$$

$$= -\int_0^\infty e^{-u}\sin \omega u du \text{ [8]}$$

$$= -\frac{\omega}{1+\omega^2}.$$

4. Given: $Q(\omega) = -\dfrac{\omega}{1+\omega^2}$. Find: $P(\omega)$ by equation (3).

$$P(\omega) = -\frac{2}{\pi}\int_0^\infty \cos \omega u du \int_0^\infty -\frac{t}{1+t^2}\sin utdt \text{ [9]}$$

$$= \int_0^\infty e^{-u}\cos \omega u du \text{ [10]} = \frac{1}{1+\omega^2}.$$

5. Given: $|Y(\omega)| = \dfrac{1}{\sqrt{(1+\omega^2)}}$. Find: $\phi(\omega)$ by equation (10).

$$\phi(\omega) = -\frac{1}{\pi}\int_0^\infty \sin \omega u du \int_{-\infty}^\infty \frac{1}{2} ln\left(\frac{1}{1+t^2}\right)\cos utdt.$$

Integrating by parts

$$\phi(\omega) = \frac{1}{2\pi}\int_0^\infty \sin \omega u du\left[\frac{1}{u} ln\left(\frac{1}{1+t^2}\right)\sin ut\Big|_{-\infty}^{\infty} - \frac{2}{u}\int_{-\infty}^{\infty}\frac{t\sin utdt}{1+t^2}\right].$$

Hence

$$\phi(\omega) = -\int_0^\infty e^{-u}\sin \omega u \frac{du}{u} \text{ [12]} = -\tan^{-1}\omega.$$

[7] Bierens de Haan, Nouvelles Tables d'Intégrales Définies, Engels, Leiden, 1867, Table 160, No. 5, p. 223.
[8] Ibid., Table 261, No. 1, p. 383.
[9] Ibid., Table 160, No. 4, p. 223.
[10] Ibid., Table 261, No. 2, p. 283.
[12] Ibid., Table 365, No. 1, p. 509.

6. Given: $\phi = -\tan^{-1}\omega$. Find: $|Y(\omega)|$ by equation (9).

$$ln|Y(\omega)| = -\frac{1}{\pi}\int_0^\infty \cos \omega u du \int_{-\infty}^\infty -\tan^{-1} t \sin ut dt.$$

Integrating by parts

$$ln|Y(\omega)| = \frac{1}{\pi}\int_0^\infty \cos \omega u du \left[-\frac{1}{u}\tan^{-1} t \cos ut \Big|_{-\infty}^{\infty} + \frac{1}{u}\int_{-\infty}^{\infty} \frac{\cos ut dt}{1+t^2} \right]^{13}$$

$$= \int_0^\infty e^{-u} \cos \omega u \frac{du}{u}.$$

Differentiating under the integral sign with respect to ω

$$\frac{d}{d\omega} ln|Y(\omega)| = -\int_0^\infty e^{-u} \sin \omega u du = -\frac{\omega}{1+\omega^2}.$$

Hence

$$ln|Y(\omega)| = -\int_0^\infty \frac{\omega d\omega}{1+\omega^2} = \frac{1}{2} ln\left(\frac{1}{1+\omega^2}\right)$$

and

$$|Y(\omega)| = \frac{1}{\sqrt{(1+\omega^2)}}.$$

7. Show that equation (16) is satisfied by $P(\omega) = \frac{1}{1+\omega^2}$, and $Q(\omega) = -\frac{\omega}{1+\omega^2}$.

Since

$$\int_{-\infty}^\infty P^2(\omega) d\omega = \int_{-\infty}^\infty \frac{d\omega}{(1+\omega^2)^2} = \int_{-\pi/2}^{\pi/2} \cos^2 x dx = \pi/2$$

and

$$\int_{-\infty}^\infty Q^2(\omega) d\omega = \int_{-\infty}^\infty \frac{\omega^2 d\omega}{(1+\omega^2)^2} = \int_{-\pi/2}^{\pi/2} \sin^2 x dx = \pi/2,$$

therefore equation (16) is satisfied.

[13] Ibid., Table 160, No. 5, p. 223.

II

It is known in electric circuit theory that the Laplace transform

$$\int_0^\infty f(x)e^{-ux}dx$$

of a function $f(x)$ vanishing to the left of the origin, is a function analytic over the right half of the complex plane, and satisfying appropriate conditions at infinity. Since the admittance function of a network is a function analytic over the right half of the complex plane, it is possible to expand this function in terms of the Laplace transforms of a complete set of functions defined for positive values of the argument. If the Laplace transforms represent the admittances of a set of known networks, and further, if it is possible to combine the admittances linearly, then the proper combination of these networks will yield a network with any desired admittance function. If we put $u = j\omega$, our Laplace transformation determines and is determined by a Fourier transformation.

An expansion of the form [14]

$$f(x) \sim \sum_{n=0}^{\infty} a_n L_n(x) \tag{17}$$

is a special Laguerre's series where $f(x)$ is defined for $x \geq 0$ and

$$L_n(x) = \frac{1}{n!} e^x \frac{d^n}{dx^n}(x^n e^{-x}), \tag{18}$$

which are called Laguerre's polynomials. The sign of equivalence ($\sim$) indicates that the coefficients are determined from the relation

$$a_n = \int_0^\infty e^{-t} L_n(t) f(t) dt. \tag{19}$$

[14] Einar Hille, "On Laguerre's Series," *Proceedings of the National Academy of Sciences*, vol. 12, 1926, pp. 261–269; and pp. 348–352.

The polynomials $L_n(x)$ have the following properties of orthogonality:

$$\int_0^\infty e^{-t}L_m(t)L_n(t)dt = \begin{cases} 0 \text{ if } m \neq n \\ 1 \text{ if } m = n. \end{cases} \tag{20}$$

If we put

$$L_n(x) = e^{x/2}h_n(x) \tag{21}$$

and

$$f(x) = e^{x/2}g(x),$$

(17) becomes

$$g(x) \sim \sum_{n=0}^{\infty} a_n h_n(x), \tag{22}$$

(18) becomes

$$h_n(x) = \frac{1}{n!}e^{x/2}\frac{d^n}{dx^n}(x^n e^{-x}), \tag{23}$$

(19) becomes

$$a_n = \int_0^\infty h_n(t)g(t)dt, \tag{24}$$

and (20) becomes

$$\int_0^\infty h_m(t)h_n(t)dt = \begin{cases} 0 \text{ if } m \neq n \\ 1 \text{ if } m = n. \end{cases} \tag{25}$$

Letting

$F(u)$ = Fourier transform of $g(x)$

and

$G_n(u)$ = Fourier transforms of $h_n(x)$,

we may write (22) as

$$F(u) \sim \sum_{n=0}^{\infty} a_n G_n(u). \tag{26}$$

By the Parseval theorem for Fourier transforms, (24) is

$$a_n = \int_{-\infty}^{\infty} F(u)\overline{G}_n(u)du. \tag{27}$$

As $G_n(u)$ are the Fourier transforms of a complete set of functions vanishing to the left of the origin, $F(u)$ may be taken to represent an admittance function. Therefore, we may write

$$Y(\omega) \sim \sum_{n=0}^{\infty} a_n G_n(\omega) \tag{28}$$

and

$$a_n = \int_{-\infty}^{\infty} Y(\omega)\overline{G}_n(\omega)d\omega. \tag{29}$$

It is evident that if $G_n(\omega)$ represent the admittances of a complete set of known networks, any desired physically realizable, quadratically summable admittance $Y(\omega)$ may be obtained by their linear combination.

To change the scale of our functions, it is only necessary to change $h_n(x)$ into $h_n(k_1x)$ where k_1 is a positive constant. This leads to

$$Y(k\mu) \sim \sum_{n=0}^{\infty} a_n G_n(\mu), \tag{30}$$

$$a_n = k\int_{-\infty}^{\infty} Y(k\mu)\overline{G}_n(\mu)d\mu, \tag{31}$$

and

$$G_n(\mu) = \frac{1}{\sqrt{k\pi}}\frac{(j\mu - 1)^n}{(j\mu + 1)^{n+1}}. \tag{32}$$

Here $\mu = \omega/k$, and $k = \dfrac{k_1}{2}$.

Before going into the physical representation of equation (30), we shall first consider the evaluation of the coefficients a_n. Separating (32) into its real and imaginary parts

$$G_n(\mu) = \frac{1}{\sqrt{k\pi}}\frac{(-1)^n}{\sqrt{(1+\mu^2)}}\cos[(2n+1)\tan^{-1}\mu] + j\frac{1}{\sqrt{k\pi}}\frac{(-1)^n}{\sqrt{(1+\mu^2)}}\sin[(2n+1)\tan^{-1}\mu]. \tag{33}$$

Letting

$$Y(k\mu) = M(k\mu) + jN(k\mu)$$

and

$$\overline{G}_n(\mu) = P_n(\mu) - jQ_n(\mu), \tag{34}$$

we have

$$\int_{-\infty}^{\infty} \overline{G}_n(\mu)\,Y(k\mu)d\mu = \int_{-\infty}^{\infty} P_n(\mu)\,M(k\mu)d\mu + \int_{-\infty}^{\infty} Q_n(\mu)N(k\mu)d\mu. \tag{35}$$

The integrals

$$\int_{-\infty}^{\infty} P_n(\mu) N(k\mu) d\mu \qquad \text{and} \qquad \int_{-\infty}^{\infty} Q_n(\mu) M(k\mu) d\mu$$

vanish since $P_n(\mu)$ and $M(k\mu)$ are even functions, and $Q_n(\mu)$ and $N(k\mu)$ are odd functions. By the Parseval theorem for Hilbert transforms which has been stated in the first part of this paper, (35) is further simplified into

$$\int_{-\infty}^{\infty} \bar{G}_n(\mu) Y(k\mu) d\mu = 2 \int_{-\infty}^{\infty} P_n(\mu) M(k\mu) d\mu$$

$$= 2 \int_{-\infty}^{\infty} Q_n(\mu) N(k\mu) d\mu. \tag{36}$$

In other words, putting

$$b_n = k \int_{-\infty}^{\infty} P_n(\mu) M(k\mu) d\mu \tag{37}$$

and

$$c_n = k \int_{-\infty}^{\infty} Q_n(\mu) N(k\mu) d\mu, \tag{38}$$

we have

$$a_n = 2b_n = 2c_n. \tag{39}$$

That is, the coefficients of (30) are completely determined by either the conductance or the susceptance of the admittance function.

Substituting the values for $P_n(\mu)$ and $Q_n(\mu)$ in (37) and (38), we have

$$b_n = k\sqrt{(k/\pi)}(-1)^n \int_{-\infty}^{\infty} \frac{\cos\left[(2n+1)\tan^{-1}\mu\right]}{\sqrt{(1+\mu^2)}} M(k\mu) d\mu, \tag{40}$$

and

$$c_n = k\sqrt{(k/\pi)}(-1)^n \int_{-\infty}^{\infty} \frac{\sin\left[(2n+1)\tan^{-1}\mu\right]}{\sqrt{(1+\mu^2)}} N(k\mu) d\mu. \tag{41}$$

Setting

$$\mu = \tan\theta, \tag{42}$$

we get

$$b_n = k\sqrt{(k/\pi)}(-1)^n \int_{-\pi/2}^{\pi/2} \frac{\cos(2n+1)\theta}{\cos\theta} M(k\tan\theta)d\theta$$

$$= k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} \left[-1 + \sum_{m=0}^{n} 2(-1)^m \cos 2m\theta \right] M(k\tan\theta)\, d\theta, \quad (43)$$

and

$$c_n = k\sqrt{(k/\pi)}(-1)^n \int_{-\pi/2}^{\pi/2} \frac{\sin(2n+1)\theta}{\cos\theta} N(k\tan\theta)d\theta$$

$$= k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} \left[\tan\theta + \sum_{m=0}^{n} 2(-1)^m \sin 2m\theta \right] N(k\tan\theta)d\theta. \quad (44)$$

The first four terms of b_n are

$$b_0 = k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} M(k\tan\theta)d\theta,$$

$$b_1 = k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} (1 - 2\cos 2\theta) M(k\tan\theta)d\theta$$

$$= b_0 - 2k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} M(k\tan\theta)\cos 2\theta d\theta,$$

$$b_2 = k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} (1 - 2\cos 2\theta + 2\cos 4\theta) M(k\tan\theta)d\theta$$

$$= b_1 + 2k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} M(k\tan\theta)\cos 4\theta d\theta, \quad (45)$$

and

$$b_3 = k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} (1 - 2\cos 2\theta + 2\cos 4\theta - 2\cos 6\theta) \times M(k\tan\theta)d\theta$$

$$= b_2 - 2k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} M(k\tan\theta)\cos 6\theta d\theta.$$

The first four terms of c_n are

$$c_0 = k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} N(k\tan\theta)\tan\theta d\theta,$$

$$c_1 = k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} (\tan\theta - 2\tan 2\theta) N(k\tan\theta)d\theta$$

$$= c_0 - 2k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} N(k\tan\theta)\sin 2\theta d\theta,$$

$$c_2 = k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} (\tan\theta - 2\sin 2\theta + 2\sin 4\theta) N(k\tan\theta) d\theta$$

$$= c_1 + 2k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} N(k\tan\theta)\sin 4\theta d\theta,$$

and

$$c_3 = k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} (\tan\theta - 2\sin 2\theta + 2\sin 4\theta - 2\sin 6\theta) \times N(k\tan\theta) d\theta$$

$$= c_2 - 2k\sqrt{(k/\pi)} \int_{-\pi/2}^{\pi/2} N(k\tan\theta)\sin 6\theta d\theta. \tag{46}$$

We thus see that the evaluation of the coefficients of the fundamental expansion (30) may be easily carried out when the preassigned characteristic of a network is graphically represented.

Our next problem is to find a complete set of networks whose admittance functions are

$$G_n(\mu) = \frac{1}{\sqrt{k\pi}} \frac{(j\mu - 1)^n}{(j\mu + 1)^{n+1}}. \tag{32}$$

These networks shall be the fundamental networks from which networks with preassigned characteristics shall be built up in accordance with equation (30).

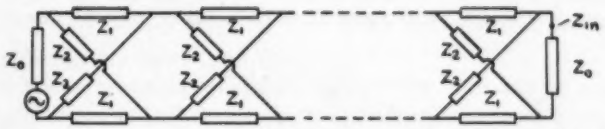


FIG. 1.

Consider the well-known lattice structure shown in Figure 1. It is terminated at both ends by equal impedances of value Z_0 which is related to the impedances Z_1 and Z_2 by the equation

$$Z_0^2 = Z_1 Z_2. \tag{47}$$

It is not difficult to show that the transfer impedance Z_{1n} at the

distant termination of a network of n sections is

$$Z_{1n} = 2\sqrt{Z_1 Z_2}\left[\frac{\sqrt{Z_2} - \sqrt{Z_1}}{\sqrt{Z_2} + \sqrt{Z_1}}\right]^n. \tag{48}$$

Now, let us shunt the distant termination by an impedance Z_s as illustrated in Fig. 2. Applying Thévenin's theorem, it may be

FIG. 2.

easily shown that the transfer impedance Z_{1p} at the shunt Z_s is

$$Z_{1p} = \frac{1/2\sqrt{Z_1 Z_2} + Z_s}{1/2\left[\frac{\sqrt{Z_2} - \sqrt{Z_1}}{\sqrt{Z_2} + \sqrt{Z_1}}\right]^n}. \tag{49}$$

Putting

$$\begin{aligned} Z_1 &= 1/j\omega C, \\ Z_2 &= j\omega L, \\ Z_s &= j\omega L/2, \\ LC &= 4, \end{aligned} \tag{50}$$

where C is a capacitance and L an inductance, (49) becomes

$$Z_{1p} = L\frac{(j\omega + 1/2)^{n+1}}{(j\omega - 1/2)^n}, \tag{51}$$

and consequently, the transfer admittance Y_{1n} at the shunt Z_s is

$$Y_{1n} = 1/Z_{1p} = 1/L\frac{(j\omega - 1/2)^n}{(j\omega + 1/2)^{n+1}}, \tag{52}$$

whose resemblance to (32) is immediately recognized. To make (52) identical with (32) it is only necessary to change the values of the network elements. Since we have made Z_1 a pure capacitive impedance and Z_2 a pure inductive impedance, Z_0 is therefore, by (47), a resistance of value $\sqrt{L/C}$.

There is a slight modification of circuit arrangements which yields the same results. Let us shunt the distant termination by a capacitance instead of an inductance. That is, let us put

$$\begin{aligned} Z_1 &= 1/j\omega C, \\ Z_2 &= j\omega L, \\ Z_3' &= 1/j\omega 2C, \\ LC &= 4. \end{aligned} \tag{53}$$

The transfer admittance Y_{1n}' at the shunted termination as shown in Fig. 3 is, by Thévenin's theorem,

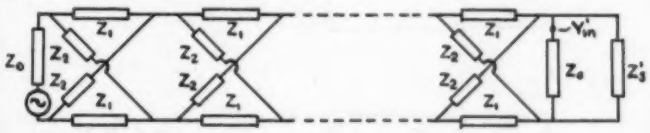


FIG. 3.

$$Y_{1n}' = 1/2L \frac{(j\omega - 1/2)^n}{(j\omega + 1/2)^{n+1}}. \tag{54}$$

This differs from (52) by only a factor of 2.

We may now form two complete sets of fundamental networks by which we synthesize networks of preassigned characteristics. Figs. 4 and 5 show the two sets of networks whose transfer admittances are $G_n(\omega)$.

It is now necessary to connect these basic networks together in such a manner that their transfer admittances combine linearly in accordance with the expansion (30). The linear combination of transfer admittances of symmetrical networks is possible. Thus in Fig. 6, if A, B, C, etc., are four-terminal networks symmetrical with respect to their horizontal axes, as indicated in the figure, and are connected to a common generator, then the short circuits connecting the output terminals of the networks, having the same voltage both in magnitude and in phase at all frequencies with reference to the generator, may be connected

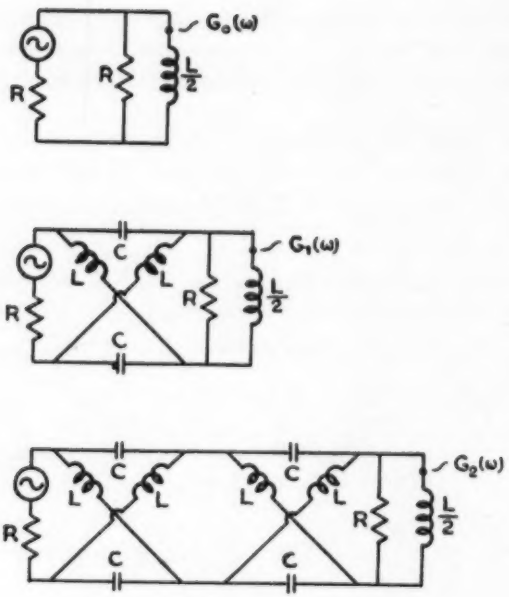


FIG. 4.

together without creating any interactions among the networks. The current at the junction S is the algebraic sum of the currents from all of the individual networks and their signs depend upon the polarity of the terminals. Hence we have here a method to combine transfer admittances linearly. In the first instance, this leads to a short-circuit transfer admittance. In order to introduce loads, we shall have to resort to the use of circuit equations. In accordance with this method of connection, the networks in Figs. 4 and 5, when composed, take the forms shown in Figs. 7 and 8 respectively.

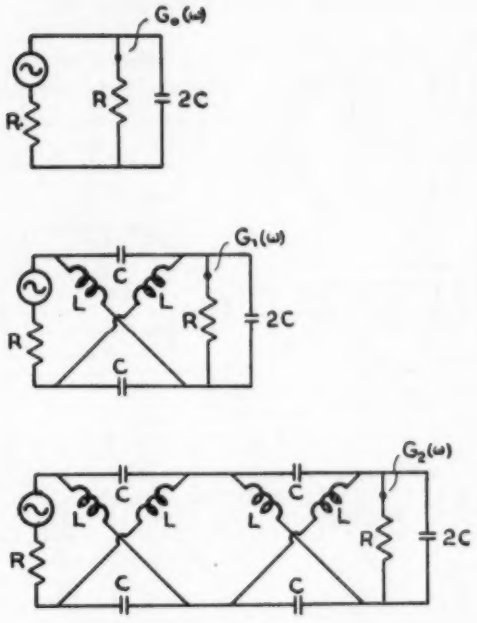

$G_0(\omega)$
R
R
2C
$G_1(\omega)$
C
L
L
R
2C
C
R
$G_2(\omega)$
C
C
L
L
L
L
R
2C
C
C
etc.

FIG. 5.

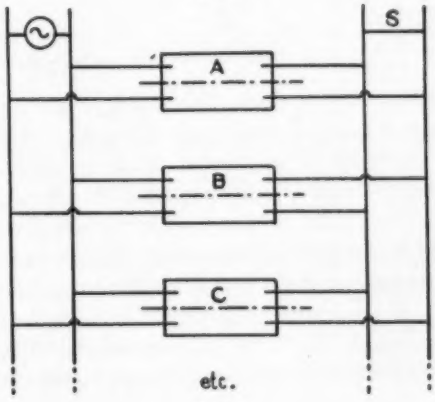

S
A
B
C
etc.

FIG. 6.

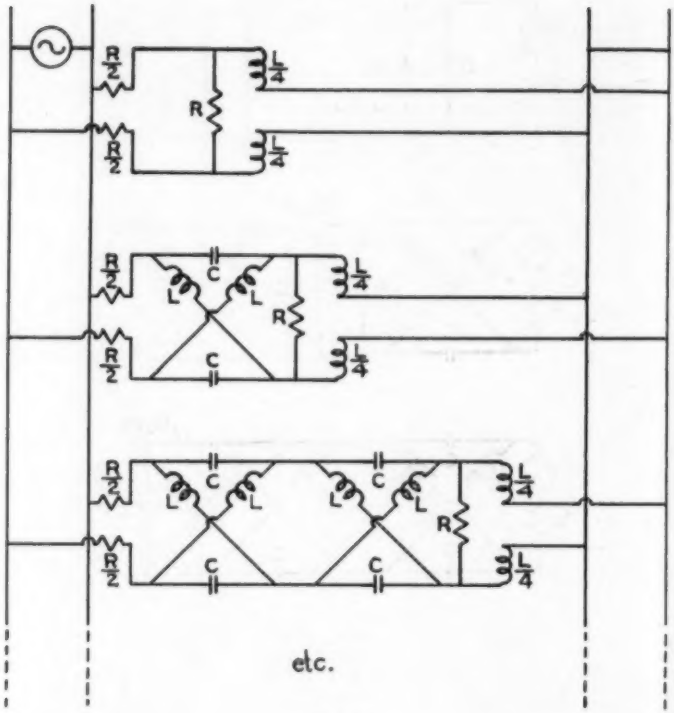


FIG. 7.

In considering loads in the synthesis of networks by the present method, we may apply a circuit theorem which is deducible from Thévenin's theorem. The theorem is: If an admittance Y be inserted into any mesh of a network, the transfer admittance of that mesh is the product of the transfer admittance prior to the addition and Y, divided by the sum of the driving-point admittance of the same mesh prior to the addition and Y. In other words, if Y_{1n} is the transfer admittance of mesh n of network J as shown in Fig. 9, Y_{nn} is the driving-point admittance of the same mesh with the generator shorted, and Y is the driving-point

admittance of the network K, then, if K is inserted into mesh n of J, the transfer admittance Y_{1n}' of mesh n is

$$Y_{1n}' = \frac{YY_{1n}}{Y + Y_{nn}}. \tag{55}$$

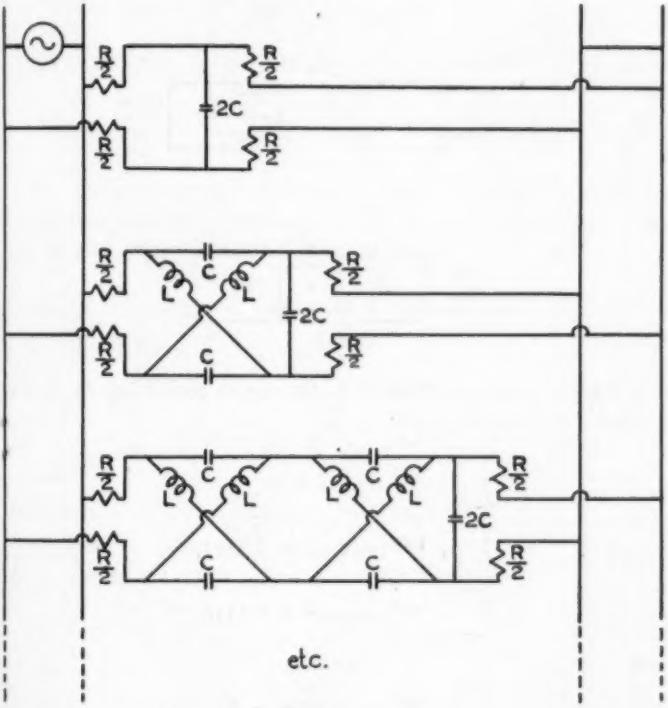


FIG. 8.

We shall now consider two simple examples. Let us specify that the conductance of the required network shall be of the form

$$e^{-\omega^2}.$$

It has already been proved that the coefficients of (30) are com-

pletely determined by either the conductance or the susceptance of the required network, and in this particular problem, it is only necessary to evaluate b_n which are simply related to a_n by (39). Referring to (45), and letting $k = 1$ for convenience, we find that

$$M(\tan\theta) = e^{-\tan^2\theta},$$

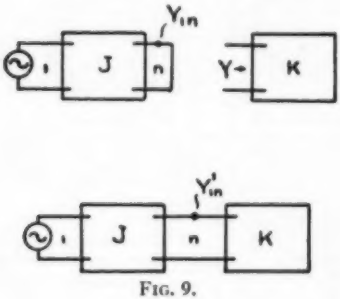


FIG. 9.

and that the first four terms of b_n are simple combinations of the following definite integrals:

$$\int_{-\pi/2}^{\pi/2} e^{-\tan^2\theta}\, d\theta = 1.3433,$$

$$\int_{-\pi/2}^{\pi/2} e^{-\tan^2\theta} \cos 2\theta d\theta = 0.8584,$$

$$\int_{-\pi/2}^{\pi/2} e^{-\tan^2\theta} \cos 4\theta d\theta = 0.1116,$$

and

$$\int_{-\pi/2}^{\pi/2} e^{-\tan^2\theta} \cos 6\theta d\theta = -0.0859.$$

According to (45)

$$2b_0 = 2/\sqrt{\pi}(1.3433) = 0.854\sqrt{\pi},$$

$$2b_1 = 2/\sqrt{\pi}[1.3433 - 2(0.8584)] = -0.2388\sqrt{\pi},$$

$$2b_2 = 2/\sqrt{\pi}[1.3433 - 2(0.8584) + 2(0.1116)] = -0.0956\sqrt{\pi}.$$

and

$$2b_3 = 2/\sqrt{\pi}[1.3433 - 2(0.8584) + 2(0.1116) + 2(0.0859)]$$
$$= 0.0137\sqrt{\pi}.$$

Putting these values in (30)

$$Y(\omega) = 0.854\frac{1}{(j\omega+1)} - 0.2388\frac{(j\omega-1)}{(j\omega+1)^2} - 0.0956\frac{(j\omega-1)^2}{(j\omega+1)^3} + 0.0137\frac{(j\omega-1)^3}{(j\omega+1)^4}\cdots. \quad (56)$$

For the purpose of checking our result, we shall evaluate the susceptance which corresponds to the given conductance by the use of Hilbert transforms. By equation (4)

$$Q(\omega) = -1/\pi\int_0^\infty \sin \omega u du \int_{-\infty}^\infty e^{-t^2}\cos utdt$$
$$= -1/\sqrt{\pi}\int_0^\infty e^{-u^2/4}\sin \omega u du$$
$$= -e^{-\omega^2}/(j2\sqrt{\pi})\int_0^\infty \cdot[e^{-(u-j2\omega)^2/4} - e^{-(u+j2\omega)^2/4}]du.$$

Letting $x = u - j2\omega$, and $y = u + j2\omega$,

$$Q(\omega) = -e^{-\omega^2}/(j2\sqrt{\pi})\left[\int_{-j2\omega}^\infty e^{-x^2/4}dx + \int_{j2\omega}^\infty e^{-y^2/4}dy\right]$$
$$= -e^{-\omega^2}/(j2\sqrt{\pi})\int_{-j2\omega}^{j2\omega} e^{-u^2/4}du$$
$$= -e^{-\omega^2}/\sqrt{\pi}\int_0^{2\omega} e^{v^2/4}dv. \quad (57)$$

This may be developed into a series by integrating by parts. In so doing we obtain

$$-\sqrt{\pi}Q(\omega) = \omega - \frac{\omega^3}{6} + \frac{\omega^5}{60} - \frac{\omega^7}{840} + \frac{\omega^9}{15120} - \frac{\omega^{11}}{332640} + \cdots.$$

For $\omega > 2$, the series

$$-\sqrt{\pi}Q(\omega) = \frac{2}{\omega} + \frac{2^2}{\omega^3} + \frac{2^3 \cdot 3}{\omega^5} + \frac{2^4 \cdot 3 \cdot 5}{\omega^7} + \frac{2^5 \cdot 3 \cdot 5 \cdot 7}{\omega^9} + \cdots$$

converges rapidly. This series is also obtained by integrating by parts.

The real and imaginary parts of (56), that is, the conductance and the susceptance of the required network are shown in Figs. 10 and 11 compared with the specified values.

We shall not go into the computation of the values of the network elements from the coefficients of (56) as the present problem is not intended for practical purposes. This step, however, is quite clear.

Suppose now we specify the susceptance of the network to be designed. Let this be

$$-\sqrt{\pi}\omega e^{-\omega^2}.$$

In this case, we shall evaluate c_n instead of b_n. Letting $k = 1$ again, we find that in (46)

$$N(\tan\theta) = -\sqrt{\pi}\tan\theta e^{-\tan^2\theta},$$

and with the definite integrals

$$\int_{-\pi/2}^{\pi/2} \tan^2\theta e^{-\tan^2\theta}\, d\theta = 0.4292,$$

$$\int_{-\pi/2}^{\pi/2} \tan\theta e^{-\tan^2\theta} \sin 2\theta d\theta = 0.4849,$$

$$\int_{-\pi/2}^{\pi/2} \tan\theta e^{-\tan^2\theta} \sin 4\theta d\theta = 0.2618,$$

and

$$\int_{-\pi/2}^{\pi/2} \tan\theta e^{-\tan^2\theta} \sin 6\theta d\theta = -0.0640,$$

the first four terms of c_n are

$$2c_0 = 2/\pi(0.4292) = 0.484\sqrt{\pi},$$

$$2c_1 = 2/\pi[0.4292 - 2(0.4849)] = -0.610\sqrt{\pi},$$

$$2c_2 = 2/\pi[0.4292 - 2(0.4849) + 2(0.2618)] = -0.0192\sqrt{\pi},$$

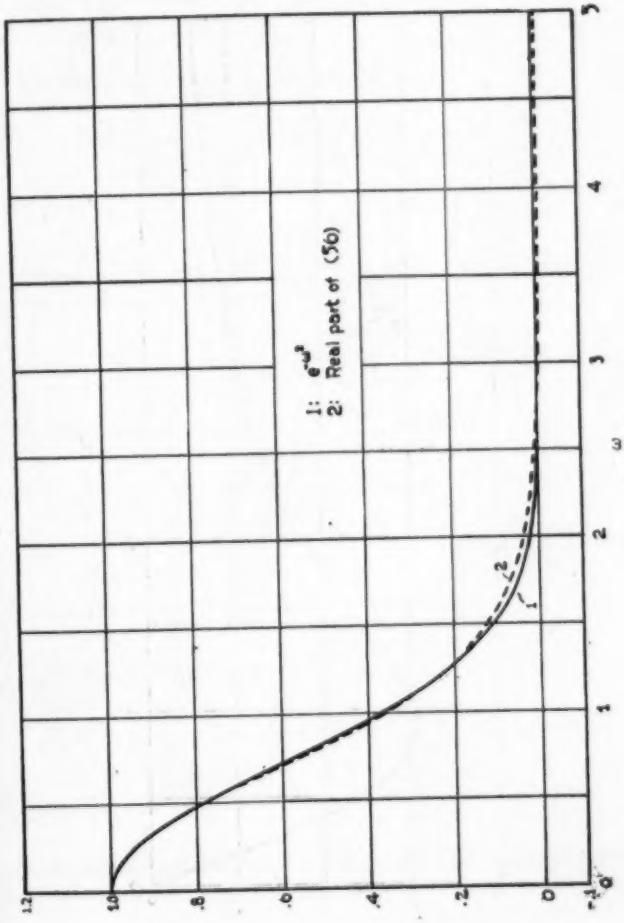


FIG. 10.

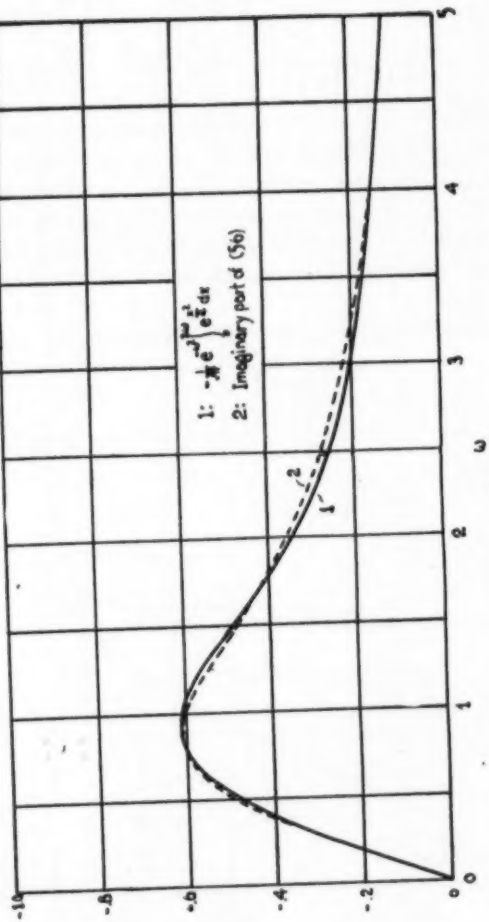

1: $-\frac{1}{\pi}e^{\omega^2/2}\int_0^\infty e^{\frac{x^3}{3}}dx$
2: Imaginary part of (5b)
-1.0
-.8
-.6
-.4
-.2
0
1
2
3
4
5
ω

FIG. 11.

and

$$2c_3 = 2/\pi[0.4292 - 2(0.4829) + 2(0.2618) + 2(0.0640)]$$
$$= 0.1253\sqrt{\pi}.$$

Putting these values in (30), we get

$$Y(\omega) = 0.484\frac{1}{(j\omega+1)} - 0.610\frac{(j\omega-1)}{(j\omega+1)^2} - 0.019\frac{(j\omega-1)^2}{(j\omega+1)^3} + 0.125\frac{(j\omega-1)^3}{(j\omega+1)^4}\cdots. \quad (58)$$

From (3) we find that the conductance corresponding to the given susceptance $-\sqrt{\pi}\omega e^{-\omega^2}$ is

$$P(\omega) = \sqrt{\pi}/\pi\int_0^\infty \cos\omega u du\int_{-\infty}^\infty te^{-t^2}\sin utdt.$$

Integrating by parts

$$P(\omega) = 1\sqrt{\pi}\int_0^\infty \cos\omega u du\left[-1/2e^{-t^2}\sin ut\Big|_{-\infty}^{\infty} + u/2\int_{-\infty}^{\infty} e^{-t^2}\cos utdt\right]$$
$$= 1/2\int_0^\infty ue^{-u^2/4}\cos\omega u du.$$

Integrating by parts again

$$P(\omega) = 1/2\left[-2e^{-u^2/4}\cos\omega u\Big|_0^\infty - 2\omega\int_0^\infty e^{-u^2/4}\sin\omega u du\right]$$
$$= 1 - \omega\int_0^\infty e^{-u^2/4}\sin\omega u du,$$

which, when compared with the expression for the susceptance in the preceding problem, is found to be

$$P(\omega) = 1 - \omega e^{-\omega^2}\int_0^{2\omega} e^{v^2/4}dv. \quad (59)$$

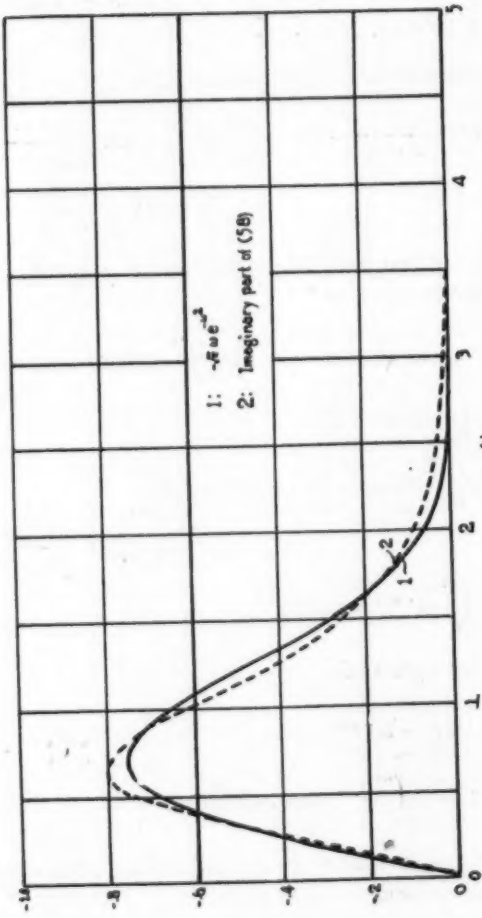

1: $-\sqrt{\pi}\,\omega\, e^{-\frac{\omega^2}{4}}$
2: Imaginary part of (58)
ω
-10
-8
-6
-4
-2
0
1
2
3
4
5

FIG. 12.

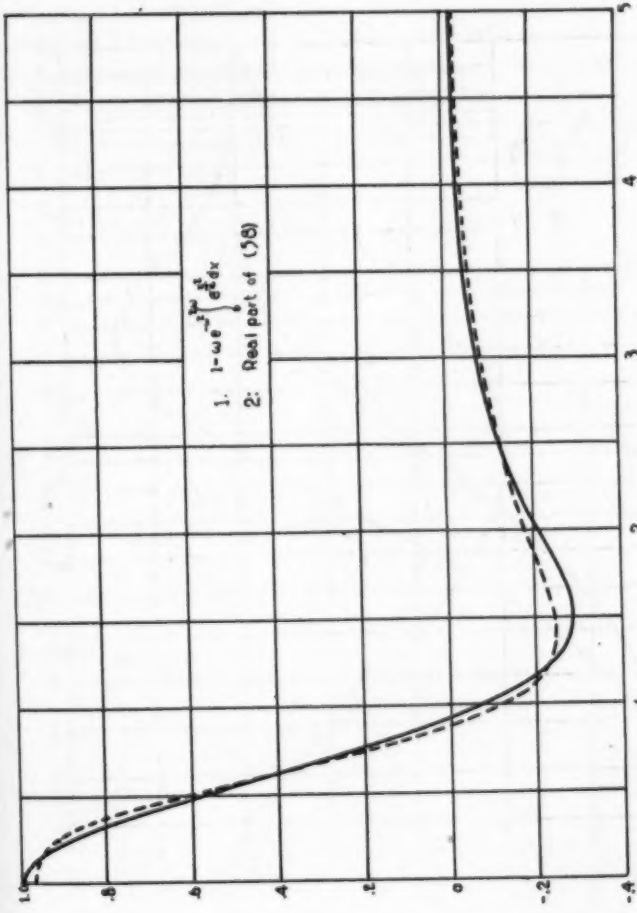

1. $1-\omega e^{-\frac{1}{2}\omega^2}\int_0^\omega e^{\frac{1}{2}x^2}dx$
2: Real part of (58)
1.0
.8
.6
.4
.2
0
-.2
-.4
0
1
2
3
4
5
ω

FIG. 13.

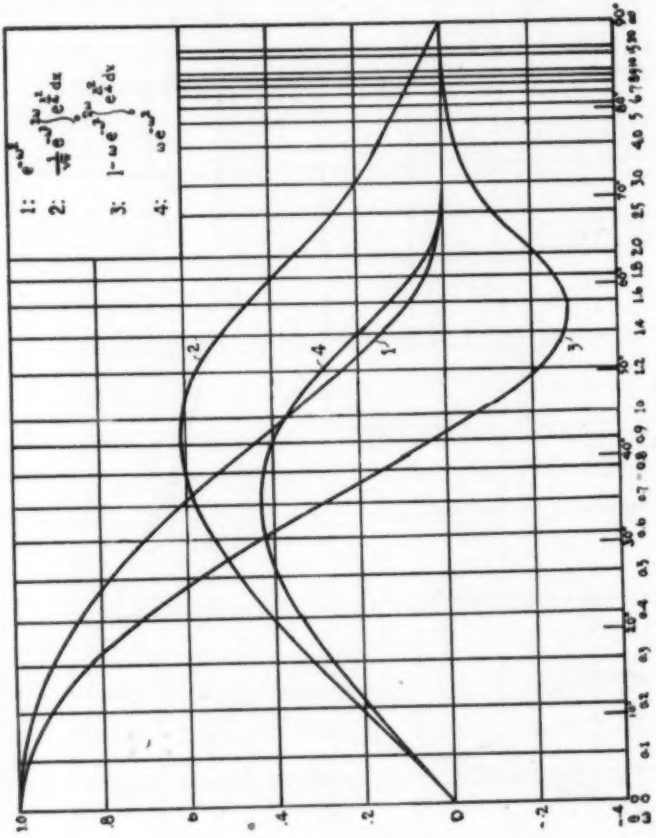

1: e^{-u^2}
2: $\frac{2}{\sqrt{\pi}}\int_0^u e^{-x^2}dx$
3: $1-ue^{-u^2}\int_0^u e^{-x^2}dx$
4: ue^{-u^2}
1'
2'
3'
4'

FIG. 14.

The real and imaginary parts (conductance and susceptance) of (58) are compared with their assigned values in Figs. 12 and 13.

Figure 14 shows $e,^{-\omega^2}$ $\omega e^{-\omega^2}$, (57), and (59) on antitangent scale computed according to (42). These graphs may be used in the computation of (45) or (46) with the aid of a harmonic analyser.

UPLIFT PRESSURE ON DAMS

By Warren Weaver *

§ 1. Introduction

When a dam rests on a porous material, such as sand or gravel, through which water can penetrate, there will be a seepage flow under the dam from the upstream side to the downstream side. Accompanying this flow there will be an upward pressure acting on the horizontal base of the dam. Such pressures have received considerable attention from design engineers. There have been important attempts to investigate, from laboratory models, the way in which this upward pressure varies across the base of the dam, and the way in which this pressure varies with head, depth of sheet-piling, etc. (See references 1 to 4.) There have been various opinions expressed, by engineers, as to the way in which this pressure varies across the base of the dam, and as to the way in which it is affected by the presence of sheet-piling. (See references 5 to 15.) There have, also, been measurements of the

* Professor Weaver was head of the department of mathematics at the University of Wisconsin until recently. He is at present with the Rockefeller Foundation. Ed.

[1] "The Action of Water under Dams," J. B. T. Coleman, Papers and Discussions, *Am. Soc. of Civil Eng.*, Vol. XLI, Aug., 1915, No. 6, p. 1295.

[2] "Designing an Earth Dam," etc., J. B. Hays, Papers and Discussions, *Am. Soc. of Civil Eng.*, Vol. XLII, March, 1916, No. 3, p. 320.

[3] "Hydrostatic Uplift in Pervious Soils," H. De B. Parsons, *Transactions, Am. Soc. of Civil Eng.*, Vol. 93, 1929, p. 1317.

[4] *Professional Memoirs, U. S. Eng. Corps*, Vol. VII, p. 44.

[5] *Uplift Pressure in Gravity Dams*, A. Floris, Western Construction News, Jan. 25, 1928, pp. 58–62.

[6] *Uplift under Dams*, C. E. Pearce, Western Construction News, Sept. 10, 1928, pp. 569–570, and Oct. 25, 1928, pp. 662–664.

[7] "J. B. Francis," *Transactions, Am. Soc. of Civil Eng.*, Vol. 91, Dec. 1927. p. 265.

[8] "O. L. Harrison," *Transactions, Am. Soc. of Civil Eng.*, Vol. LXXV, pp, 142–225.

existing pressures under actual dams. (See references 16 to 20.)

Certain statements, assumptions, or procedures noted in these papers are here listed, numbering them for ease of reference. (1) It is often assumed that the uplift pressure varies linearly from the full head-water pressure at the heel of the dam to zero pressure at the toe. (2) It is often assumed that there is, in the case of non-zero tail-water pressure, a "trapezoidal" distribution of pressure; that is, a pressure which drops linearly from the head-water pressure at the heel to the tail-water pressure at the toe. (3) It is sometimes assumed that the pressure varies linearly, but that the pressure at the heel is some proper fraction of the head-water pressure, this reduction sometimes being the method used to take account of the particular good or bad qualities of a given foundation which is partly permeable and partly impervious rock. (4) An alternative method of taking account of such conditions is to assume that the uplift pressures act only on some fractional

[9] *Masonry Dam Design, Including High Masonry Dams*, C. E. Morrison and O. L. Brodie, New York, 1916, pp. 12–14.

[10] *Engineering for Masonry Dams*, W. P. Creager, New York, 1917, p. 32.

[11] *Dams*, in *Hool and Johnson's Concrete Eng. Handbook*, A. G. Hillberg, New York, 1913, p. 731.

[12] *Irrigation Engineering*, A. P. Davis and H. M. Wilson, New York, 1919, p. 45.

[13] "Notes on Arched Gravity Dams," *Proceedings, Am. Soc. of Civil Eng.*, Aug., 1927, p. 1140.

[14] "H. Fecht," *Zentralblatt der Bauwesen*, 1889, pp. 397 and 443.

[15] "Talsperren und Wasserkraftanlangen," in Förster's *Taschenbuch für Bauinginieure*, Berlin, 1923, p. 1231.

[16] "Upward Pressures under Dams: Experiments by the U. S. Bureau of Reclamation," J. Hinds, *Transactions, Am. Soc. of Civil Eng.*, Vol. 93, 1929, pp. 1527–1550; also discussion by I. E. Houk, same volume, pp. 1565–1571.

[17] "German Studies of Uplift on Masonry Dams," A. W. Hoffman, *Engineering and Contracting*, Sept. 22, 1915, pp. 218–221.

[18] "Unterdruck bei Stamauren," R. Schaefer, *Zeitschrift für Bauwesen*, 1913, p. 101.

[19] "Experiments on Uplift in Masonry Dams," C. R. Weidner, *Eng. News*, July 31, 1913, pp. 202–205.

[20] "Tests of Water Pressure under Brule River Dam," D. W. Mead and C. V. Seastone, *Engineering News-Record*, Feb. 18, 1926, pp. 274–275.

portion of the base. (5) It is sometimes assumed that the uplift pressure is uniform across the base, and is equal to the head-water pressure or some proper fraction of it. (6) Suppose that the base of a dam is $b = b_1 + b_2$ feet wide, and that sheet-piling of depth d feet is located b_1 feet from the heel of the dam. It has been assumed that the uplift pressures which operate on the b_1 feet of base upstream from the piling are those which would operate on the first (upstream) b_1 feet of a fictitious dam $b_1 + 2d + b_2$ feet wide; while the pressures downstream from the piling are those which would operate on the last (downstream) b_2 feet of the base of the fictitious dam. This is referred to as the "line of creep" theory, and the assumption is based on the notion that the pressure drops linearly as the moisture "creeps" along the base of the dam, down one face of the sheet-piling, up the other face, and then again along the base of the dam. (7) Approximate analyses have been based on the idea that the lines of seepage flow through the sand or soil are straight, one author supposing that these straight lines make an angle of 45° with the horizontal, down under the sheet-piling and then back up again. (8) Another author supposes that the flow proceeds on straight lines from the bottom of the sheet-piling to the toe of the dam. He then calculates, according to an arbitrary assumption, the pressure along this line, and determines the depth that the sheet-piling must have in order that the total force due to this pressure equal the weight of the triangular wedge of sand or soil contained between the base of the dam, the sheet-piling, and the line, just mentioned, from the bottom of the sheet-piling to the toe of the dam. This author apparently believes that with sheet-piling of this depth there will be no uplift pressure on the base of the dam. (9) Some engineers have urged that the depth of sheet-piling be at least equal to the maximum head at flood-stage. (10) None of the papers which the author has consulted contains or refers to any attempt to solve the problem accurately on the basis of the fundamental laws for the flow of underground waters.

There are many practical and technical considerations that

enter into the general problem of uplift pressures and their influence on dam design which require engineering judgment and engineering experience. The present author is in no way qualified to deal with such considerations. This paper presents an exact solution of the problem of uplift pressures under the highly idealized condition of uniform foundation material. The results of this paper must therefore be interpreted as an idealized but basic theory which should be modified to meet practical circumstances. There is, however, a real point in knowing the underlying exact theory, since it provides a basis for intelligent modification, and since the exact theory is capable of furnishing a definite answer to most of the questions raised by the numbered remarks of the preceding paragraph.

§ 2. The Mathematical Problem

Figure 1 represents the cross-section of a gravity dam whose base measures b feet, there being a head h_0 of water upstream, and, in the simplest instance, a zero head downstream from the toe of

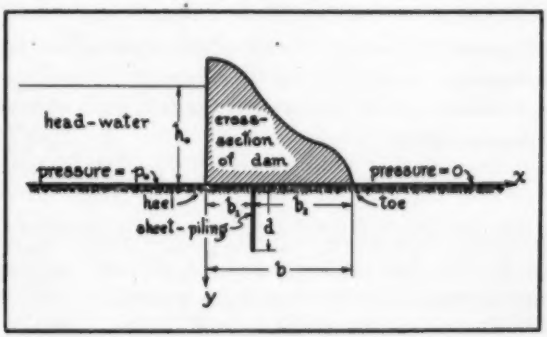


Fig. 1

the dam. At a distance b_1 from the heel, sheet-piling extends downward to a depth d, the distance of this piling from the toe being denoted by $b_2 = b - b_1$. This study will suppose the dimension of the dam perpendicular to the section shown to be

sufficiently large, and the conditions in this perpendicular direc tion sufficiently uniform so that one may treat the problem as a two-dimensional one. One chooses a system of rectangular coordinates whose origin is located at the heel of the dam, whose positive x-axis extends horizontally to the right, and whose positive y-axis extends vertically down. (See Fig. 1.)

In 1856 Henri Darcy [21] first stated the fundamental law for the motion of undergfound waters, namely, the law that the velocity of seepage flow is proportional to the pressure gradient. Much important analysis [22] has related to the effect of size of grain, density and type of packing, temperature, etc., on the proportionality factor of this law; but in the present problem this factor, as will be seen very shortly, cancels out and hence is of no importance. If $p(x, y)$ represent the pressure at any point x, y one has, by the law quoted above,

$$(1) \qquad v_x = k\frac{\partial p}{\partial x}, \qquad v_y = k\frac{\partial p}{\partial y},$$

where k is the Darcy constant, and where v_x and v_y are the x and y components, respectively, of the velocity of seepage flow. Since, moreover, one is dealing with a steady state of flow of an incompressible fluid,

$$(2) \qquad \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{dy} = 0,$$

this so-called "equation of continuity" expressing the fact that there is zero net flow into or out of a small element of volume. If one substitutes from (1) into (2) the result is

$$\frac{\partial}{\partial x}\left(k\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial y}\left(k\frac{\partial p}{\partial y}\right) = 0,$$

[21] *Les Fontaines Publiques de la Ville de Dijon*, H. Darcy, 1856, p. 590.

[22] See *Grundwasserbewegung*, Ph. Forcheimer, Encyklopädie der Mathematischen Wissenschaften, Band IV, 3, Art. 20, V, pp. 445–458, for general references. The basic analysis was made by C. S. Slichter, Nineteenth Annual Report, U. S. Geological Survey, Part II.

or, since the material is assumed homogeneous so that k does not depend on x or y,

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} = 0. \tag{3}$$

Thus the pressure satisfies the equation of Laplace. To complete the analytical formulation of the problem it remains only to specify the boundary conditions on the function $p(x, y)$. On the negative x-axis (that is, under the head-water) the pressure obviously must be constant. That is,

$$p(x, y) = p_0 = \frac{h_0}{\rho}, \quad \text{for} \quad y = 0, \quad x \leq 0, \tag{4}$$

where ρ is the density of the head-water. To the right of the toe of the dam, the pressure is also constant, but zero. That is,

$$p(x, y) = 0, \quad \text{for} \quad y = 0, \quad x \geq 0. \tag{5}$$

The lines of seepage flow will be everywhere normal to the curves $p(x, y) =$ constant, and these lines of flow must therefore be normal to the x-axis for $x \leq 0$ and for $x \geq b$, while the lines of flow must be tangential to the upstream and downstream faces of the piling, as well as to the base of the dam. It is interesting to determine these lines of seepage flow, and to know the pressure p at all points x, y; but the essential information sought is the value $p(x, 0)$, $0 \leq x \leq b$, of the uplift pressure acting on the base of the dam.[23]

[23] It is essential that the analytic problem be "well set"; that is, it is essential that there exist one and only one function which satisfies the differential equation and the boundary conditions. Since p satisfies the Laplace equation, and since the curves of flow, say $u =$ constant, and the equi-pressure curves $p =$ constant are orthogonal, it follows that $u(x, y)$ is a solution of the Laplace equation, and that u is, indeed, the function conjugate to $p(x, y)$, in the sense that $u(x, y) + ip(x, y)$ be an analytic function of the complex variable $z = x + iy$. The conditions on the function $u(x, y)$ are such as to insure the uniqueness of the solution in the region $y \leq 0$ bounded above by $y = 0$, $x \leq 0$; by $x = 0$, $y \leq d$; and by $y = 0$, $x \geq 0$. In fact, one specifies that the normal derivative of u vanish for $x \leq 0$ and for $x \geq b$; that the tangential derivative vanish on both sides of the sheet-piling and on the base of the dam;

§ 3. SOLUTION IN THE CASE OF NO SHEET-PILING [24]

It is well known that if

$$w = u + iv = f(z) = f(x + iy)$$

be an analytic function of the complex variable $z = x + iy$, then the conjugate functions $u(x, y)$ and $v(x, y)$ are solutions of the Laplace equation, and the curves $u =$ constant and $v =$ constant form orthogonal families. Consider the analytic function,

$$w = \cosh^{-1} \frac{(2z - b)}{b}, \tag{6}$$

or

$$2z - b = b \cosh w.$$

Then

$$\begin{aligned} 2x - b + i2y &= b \cosh (u + iv) \\ &= b (\cosh u \cos v + i \sinh u \sin v), \end{aligned}$$

so that

$$b \cosh u \cos v = 2x - b, \tag{7}$$

$$b \sinh u \sin v = 2y, \tag{8}$$

---

and that the behavior of u at more and more distant points approach the function characteristic of this same problem if the dam and the sheet-piling be absent. That is, the very distant lines of flow must approach semi-circles with center at the origin. Since the problem admits but one solution for u, and since the associated conjugate function p is determined, except for a constant, by u, it is clear that the solution for p is unique except for a constant. It will be seen that this constant is of physical importance.

It is clear that the boundary conditions (4) and (5) on p do not, together with (3), serve to determine p uniquely; for these conditions do not distinguish between a case of no sheet-piling and a case in which sheet-piling is present. It is, to be sure, possible to assign on p further boundary conditions which bear on the base of the dam and the two sides of the sheet-piling. Here, in fact, the normal derivative of p must vanish. The only advantage of centering attention, in connection with the question of uniqueness, on the function u rather than on p is that it is somewhat easier to assign proper behavior at infinity to u.

[24] The solution given in this section is not original with the author, but is taken from notes written in 1915 under Professor Henry Wolff, now of Drexel Institute.

and

$$\frac{(2x-b)^2}{b^2\cos^2 v} - \frac{(2y)^2}{b^2\sin^2 v} = 1, \tag{9}$$

$$\frac{(2x-b)^2}{b^2\cosh^2 u} + \frac{(2y)^2}{b^2\sinh^2 u} = 1. \tag{10}$$

The quantities u and v, defined as functions of x and y by these last two equations, are, by virtue of their origin in the analytic relation (6), solutions of the Laplace equation. Moreover, the curves u = constant, as is evident from (10), are confocal ellipses with origin at the left focus, the distance between foci being b. Thus the curves u = constant meet the x-axis normally when $x \leq 0$ or when $x \geq b$, and are tangential to the x-axis for $0 \leq x \leq b$. Moreover, the curves u = constant approach, for points distant from the origin, the form of circles about the origin. Thus (see footnote 23) the curves u = constant depict the lines of seepage flow under a dam with no sheet-piling. The conjugate function $v(x, y)$ is therefore, except for a possible additive constant, equal to the pressure function. Now when $v = 0$ it follows from (7) and (8) that $y = 0$ while $x \geq b$; while $v = \pi$ corresponds to $y = 0$, $x \leq 0$. Thus $v(x, y)$ is the pressure function for the case of a head-water pressure of π units and a tail-water pressure of zero. The modification for the case, in hand, of a head-water pressure p_0 requires merely a change of pressure scale. Thus

$$p(x, y) = \frac{p_0}{\pi} v(x, y) \tag{11}$$

is the pressure function sought, and the lines of equal pressure are the confocal hyperbolas (9). When $u = 0$, it follows from (7) and (8) that $y = 0$, $0 \leq x \leq b$. Thus $u = 0$ across the base of the dam where the uplift pressure is sought. This uplift pressure is hence given, by (7), as

$$p(x, 0) = \left[\frac{p_0}{\pi} v(x, y)\right]_{u=0} = \frac{p_0}{\pi} \cos^{-1} \frac{2x-b}{b}. \tag{12}$$

The discussion of the preceding two paragraphs is direct, but does not indicate how one might originally have chosen the function (6) which furnishes the solution. The conformal mapping, however, of the w-plane on the z-plane by the function $z = \cosh w$ is well known, and may be described geometrically as follows: Consider the horizontal strip of the w-plane between the lines $v = 0$ and $v = \pi$, and think of these lines as being broken and pivoted at the points where $u = 0$. Rotate the strip 90° counterclockwise, and at the same time fold each of the broken lines $v = 0$ and $v = \pi$ back on itself, the strip thus being doubly "fanned out" so as to cover the entire z-plane.

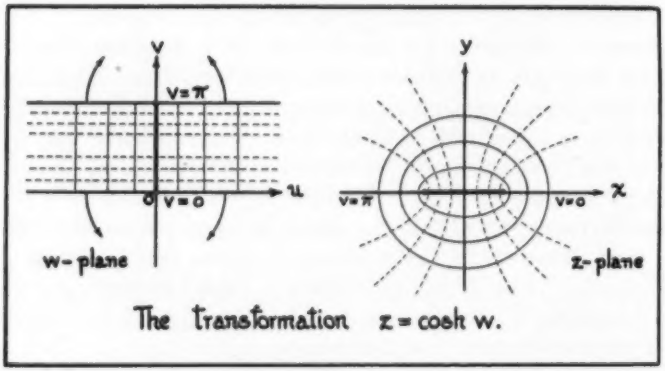


FIG. 2

This plane transformation is shown in Fig. 2, where the full lines in the z-plane are the lines of seepage flow, and the dotted lines, the lines of equal pressure. Familiarity with this transformation leads one to select the form of equation (6), and modifications of the problem, to be introduced presently, are to be handled in a similar way from one's knowledge of conformal representation.

§ 4. Discussion of Results: No Sheet-Piling

The distribution of pressure across the base of a dam without sheet-piling is given by (12) and is shown in Fig. 3. It will be noted that the frequent assumption that the pressure drops

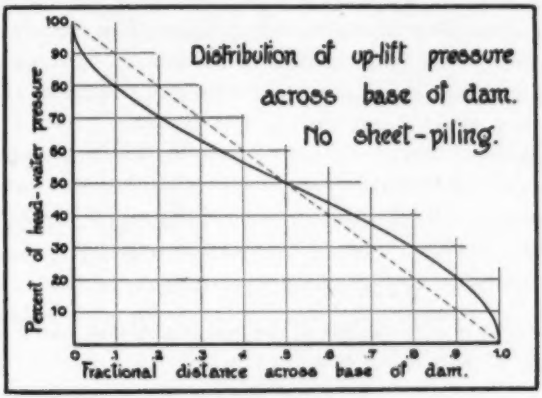


Fig. 3

linearly is not rigorously justified, but nevertheless the assumption is not far from true. The straight line assumption assigns too great pressures to the upstream half of the dam, and too small pressures to the downstream half. The total upward force F, acting per foot of dam measured perpendicular to the section being studied, is

$$F = \frac{p_0}{\pi}\int_0^b \cos^{-1}\frac{2x-b}{b}\,dx, \tag{13}$$
$$= \frac{p_0 b}{2},$$

which is exactly the value obtained on the straight line assumption. The moment M of the uplift forces relative to the heel of the dam is given by

$$M = \frac{p_0}{\pi}\int_0^b x\cos^{-1}\frac{2x-b}{b}\,dx, \tag{14}$$
$$= \frac{3b^2 p_0}{16}.$$

Hence the moment arm X of the total uplift F is

$$X = \frac{3b^2p_0}{16} \cdot \frac{2}{p_0b} = \frac{3b}{8}. \tag{15}$$

On the basis of the straight line assumption the moment is $p_0b^2/6$ and the point of application of the resultant upward thrust is located $b/3$ feet from the heel of the dam. The straight line assumption thus results in an uplift moment, relative to the heel of the dam, which is too small by 11.1 per cent.

It has been noted above that the function p is, as derived from the uniquely determined function u, determined only to an additive constant. If one includes this additive constant, and if one provides also for a change of pressure scale, one has

$$p(x, y) = k_1v(x, y) + k_2.$$

Since $v(x, y) = \pi$ to the left of the heel of the dam, and $v = 0$ to the right of the toe, one has

$$p_0 = k_1\pi + k_2,$$
$$p_1 = k_2,$$

where p_0 and p_1 are the head and tail-water pressures respectively. Thus, in general,

$$p(x, y) = \frac{p_0 - p_1}{\pi} v(x, y) + p_1. \tag{12'}$$

The pressure thus drops from the head-water value to the tail-water value along the inverse cosine curve shown in Fig. 3. For this general case of non-vanishing tail-water pressure one has, corresponding to (13), (14), and (15), the values:

$$F = \int_0^b \left[\frac{p_0 - p_1}{\pi} \cos^{-1} \frac{2x - b}{b} + p_1 \right] dx = \frac{b(p_0 + p_1)}{2}, \tag{13'}$$

$$M = \int_0^b x \left[\frac{p_0 - p_1}{\pi} \cos^{-1} \frac{2x - b}{2} + p_1 \right] dx = \frac{b^2(3p_0 + 5p_1)}{16}, \tag{14'}$$

$$(15') \qquad X = \frac{b^2(3p_0 + 5p_1)}{16} \cdot \frac{2}{b(p_0 + p_1)} = \frac{b(3p_0 + 5p_1)}{8(p_0 + p_1)},$$

while the trapezoidal law described in § 1, 2 assigns values:

$$F = \int_0^b \left[p_0 - \frac{p_0 - p_1}{b} x \right] dx = \frac{b(p_0 + p_1)}{2},$$

$$M = \int_0^b x \left[p_0 - \frac{p_0 - p_1}{b} x \right] dx = \frac{b^2(p_0 + 2p_1)}{6},$$

$$X = \frac{b(p_0 + 2p_1)}{3(p_0 + p_1)}.$$

From these expressions it is observed that the trapezoidal assumption produces the correct total uplift; while it is easily calculated that the per cent error in the moment assigned by the trapezoidal assumption is given by the expression,

$$100 \frac{1 - l}{9 + 15l},$$

where $l = p_1/p_0$. This per cent error for various ratios l of p_1 to p_0 is shown in the following brief table:

TABLE I

THE PER CENT ERROR IN THE UPLIFT MOMENT AS CALCULATED FROM THE TRAPEZOIDAL ASSUMPTION FOR VARIOUS RATIOS l OF TAIL-WATER PRESSURE TO HEAD-WATER PRESSURE

l	Per cent error
0.0	11.1
0.1	8.6
0.2	6.7
0.3	5.2
0.4	4.0
0.5	3.0
1.0	0.0

§ 5. SOLUTION WHEN SHEET-PILING IS PRESENT

To find the lines of equal pressure and the lines of flow when sheet-piling is present, it will be necessary to use a second analytic transformation, in addition to that expressed by (6). Since it

will be convenient to continue to use x and y as the independent coordinates which locate points relative to the head of the dam, transformation (6) will be rewritten as

$$w = \cosh^{-1} \frac{2\zeta - b'}{b'}. \tag{6'}$$

This maps the ζ-plane on the w-plane.[25] The second transformation, just referred to, will map the ζ-plane on a z-plane, and the combination of the two transformations will map the w-plane directly onto the z-plane, the intermediate variable ζ being thus eliminated.

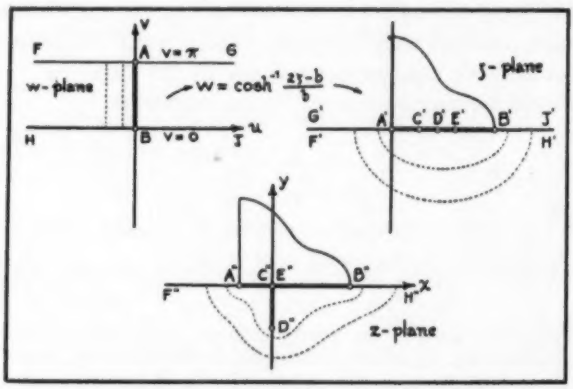


FIG. 4

The conformal transformation (6') maps, as indicated in Fig. 4, the (dotted) vertical straight lines u = constant of the w-plane into the ellipsoidal lines of flow in the ζ-plane. The points lettered on the w-plane transform into the corresponding primed points of the ζ-plane. When sheet-piling is present, the lines of flow must continue to meet the x-axis (see the z-plane of Fig. 4)

[25] The letter b has been used to indicate the width of the base of the dam. The constant b of (6) is written as b' in (6') for the reason that, after a second transformation has been applied, the base of the dam will have a new width which we wish to continue to call b.

normally to the left of the heel A'' and to the right of the toe B'' of the dam, and must continue to be tangential to the base of the dam. In addition, however, the lines of flow must now also be tangential to the two faces of the sheet-piling. Consider a conformal transformation which transforms the lower half ζ-plane into the lower half z-plane, the axis of reals in the ζ-plane going into the x-axis of the z-plane except that a portion $C'D'E'$ of the real axis contained entirely within the base of the dam is transformed into the vertical line segment $x = 0$, $y \leq b$ of the z-plane. That is, suppose the point D' is, by the transformation, displaced downward to D'', while the point C' is displaced to the right and E' to the left until they coincide at the point C'' (or E'') directly over D''. In such a conformal transformation any line of flow will, in its displaced configuration, continue to intersect the x-axis normally, and the displaced lines of flow will clearly be tangential to both faces of the sheet-piling as well as to the base of the dam.

It is easy to obtain the conformal transformation just described, since the lower half z-plane bounded above by $F''A''C''D''E''B''H''$ constitutes a rectilinear polygon which may be mapped on the entire lower half ζ-plane by means of a Schwartz-Christoffel transformation.[26] The Schwartz-Christoffel theorem states that the interior of a rectilinear polygon in a complex z-plane can be conformly mapped onto the entire positive half ζ-plane by means of the transformation,

$$z = A \int \frac{d\zeta}{(\zeta - \zeta_1)^{\beta_1}(\zeta - \zeta_2)^{\beta_2} \cdots (\zeta - \zeta_n)^{\beta_n}} + B,$$

where the sides of the polygon, when one traverses the boundary in the positive sense, change direction by an angle $\beta_i\pi$ when the i'th vertex is passed, and where ζ_i is the point on the axis of reals of the ζ-plane into which the i'th vertex maps. The constants A and B determine the size and orientation of the polygon in the

[26] *Ueber einige Abbildungsaufgaben*, Crelle 70, 1869, pp. 105–120; "Sul problema delle Temperature stazionarie," *Annali di Matematica*, I, 1867, p. 89.

ζ-plane. In the case in hand, the polygon has a vertex $H''E''D''$ at E'' for which $\beta = 1/2$, a vertex $E''D''C''$ at D'' for which $\beta = -1$, and a vertex $D''C''F''$ at C'' for which $\beta = 1/2$. One may conveniently choose to map E'' on $\zeta = +1$, D'' on $\zeta = 0$, and C'' on $\zeta = -1$. The transformation is, then, writing for the moment $A = 1$, $B = 0$,

$$z = \int \frac{\zeta d\zeta}{\sqrt{\zeta^2 - 1}} = \sqrt{\zeta^2 - 1},$$

or

$$\zeta = \sqrt{1 + z^2}.$$

It is readily checked that with this choice of the constants A and B the distance $D''E''$ is of unit length, while C'' and E'' are located at the origin of the z-plane. If one wishes $D''E''$ to have the length d (the depth of the sheet-piling); C'' and E'' to remain at $x = y = 0$; and the distances $A'D'$ and $C'D'$ to have the values c and e respectively, one writes, by mere translations and changes of scale,

$$\frac{\zeta - c}{e} = \sqrt{1 + \left(\frac{z}{d}\right)^2}. \tag{16}$$

In light of the above remarks, it follows that, when sheet-piling is present, the stream lines and the lines of equal pressure are obtained by setting constant the real and pure imaginary parts u and v of w, where

$$\frac{b' \cosh w + b'}{2e} - \frac{c}{e} = \sqrt{1 + \left(\frac{z}{d}\right)^2},$$

or

$$\lambda \cosh w + \lambda - \varkappa = \sqrt{1 + \left(\frac{z}{d}\right)^2}, \tag{17}$$

where

$$\lambda = \frac{b'}{2e},$$

$$\varkappa = \frac{c}{e}; \tag{18}$$

the transformation (17) resulting from applying (6′) and then (16). One should note that both λ and $\varkappa$ are greater than or equal to 1.

It is necessary to know the abscissas of the points A'' and B'' so that we can calculate the parameters λ and $\varkappa$ in terms of the base width b and the distance b_1 from the heel to the sheet-piling. The point A'' is the map of the point A, for which $u = 0, v = \pi$. Thus the abscissa of A'' is given by the negative value of z in

$$\lambda \cosh (0 + i\pi) + \lambda - \varkappa = \sqrt{1 + \left(\frac{z}{d}\right)^2},$$

or

$$-\varkappa = \sqrt{1 + \left(\frac{z}{d}\right)^2}.$$

Thus,

$$z_{A''} = -d\sqrt{\varkappa^2 - 1} = -b_1. \tag{19}$$

Similarly, for B'',

$$\lambda \cosh (0 + i \cdot 0) + \lambda - \varkappa = \sqrt{1 + \left(\frac{z}{d}\right)^2},$$

$$z_{B''} = +d\sqrt{(2\lambda - \varkappa)^2 - 1} = b_2. \tag{20}$$

The general solution of the case in which sheet-piling is present and is located at any distance b_1 from the heel of the dam is furnished by the transformation (17), the parameters (18) being calculated from (19) and (20)

§ 6. Discussion: Sheet-Piling at the Heel of the Dam

When the sheet-piling is located at the heel, $b_1 = 0$, so that $\varkappa = 1$. Moreover,

$$+ d\sqrt{(2\lambda - \varkappa)^2 - 1} = + 2d\sqrt{\lambda^2 - \lambda} = b_2 = b_1$$

or

$$\lambda = \frac{1 + \sqrt{1 + \alpha^2}}{2}, \tag{21}$$

where

$$\alpha = \frac{b}{d} \tag{22}$$

is the ratio of width of base to the depth of the sheet-piling, and where the choice of sign of the radical in (22) is dictated by the fact, noted above, that λ is greater than or equal to 1. Inserting the value $\kappa = 1$ into (17) and equating reals and pure imaginaries,[27]

$$\lambda \cosh u \cos v + \lambda - 1 = \pm \sqrt{\frac{R + \sqrt{R^2 + 4S^2}}{2}}, \qquad \begin{matrix} + \text{ for } x \geq 0 \\ - \text{ for } x \leq 0, \end{matrix}$$

$$\lambda \sinh u \sin v = - \sqrt{\frac{-R + \sqrt{R^2 + 4S^2}}{2}}, \tag{23}$$

where

$$R = \frac{d^2 + x^2 - y^2}{d^2},$$

$$S = \frac{xy}{d^2}. \tag{24}$$

Along the base of the dam $u = 0$, $0 \leq x \leq b$, and $y = 0$, so that the first equation (23) gives

$$\cos v = \frac{d - \lambda d + \sqrt{d^2 + x^2}}{\lambda d};$$

or, inserting the scale factor for pressure,

$$p(x, y) = \frac{p_0}{\pi} v(x, y) = \frac{p_0}{\pi} \cos^{-1}\left[\frac{d - \lambda d + \sqrt{d^2 + x^2}}{\lambda d}\right], \tag{25}$$

the parameter λ being given, in terms of the ratio α of base width to piling depth by (21). In Fig. 5 is shown the distribution of uplift pressures across the base of dams with various depths of sheet-piling driven at the heel. The vertical scale represents per cent of p_0, the head-water pressure, while the horizontal scale

[27] The signs of the radicals in equations (23) result from a detailed study of the transformation (16). One may easily check that the indicated signs choose the branches of the multiple-valued function which correspond to the geometrical description, given earlier, of the desired transformation.

measures fractional parts of b, the width of the base. The lowest curve corresponds to a case in which the sheet-piling is as deep as the base is wide; the piling, in the case of the next to highest

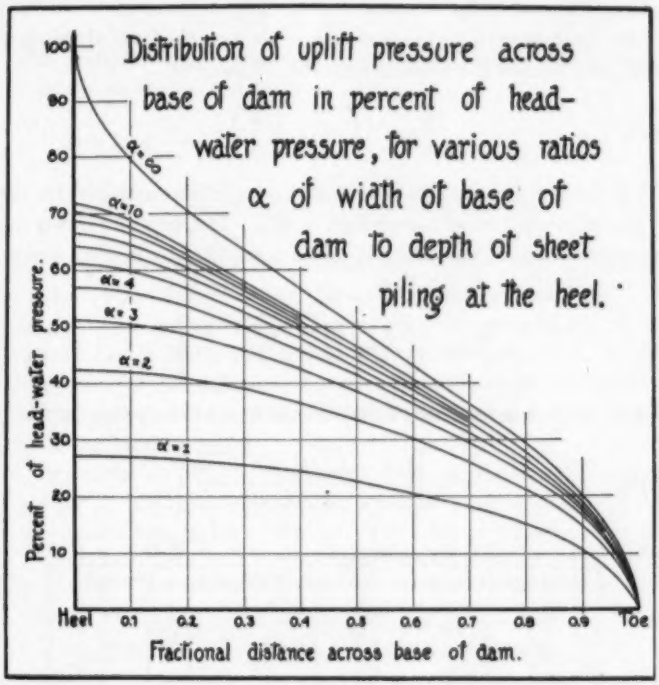


FIG. 5

curve, is one-tenth as deep as the base is wide; while the uppermost curve corresponds to no sheet-piling. One observes from the curves that the linear assumption of pressure distribution is less and less accurate, the greater the depth of the sheet-piling. With no sheet-piling, for example, the linear assumption assigns the correct pressure at the mid-point of the base, while, in the case $\alpha = 1$, the error, at the mid-point, in the linear assumption is

obviously large. One cannot say just what this error is until he agrees what pressure he assumes at the heel. This question will, therefore, be discussed later in connection with the consideration of the "line of creep" theory.

In the protected corner $y = 0, x = +0$ on the base of the dam and just behind the sheet-piling, the pressure is

$$\frac{p_0}{\pi}\cos^{-1}\left(-1+\frac{2}{\lambda}\right). \tag{26}$$

This value will presently be compared with that assigned by the "line of creep" theory described in § 1. The pressure down the upstream face of the piling is found from (23) by setting $u = 0$, $x = -0, 0 \leq y \leq d$. It is

$$\frac{p_0}{\pi}\cos^{-1}\left[\frac{d-\lambda d-\sqrt{d^2-y^2}}{\lambda d}\right]; \tag{27}$$

while the pressure on the downstream face of the piling is

$$\frac{p_0}{\pi}\cos^{-1}\left[\frac{d-\lambda d+\sqrt{d^2-y^2}}{\lambda d}\right], \tag{28}$$

obtained from (23) by setting $u = 0$, $x = +0$, $0 \leq y \leq d$.

The total uplift force on the base of the dam is given by

$$\begin{aligned} F &= \frac{p_0}{\pi}\int_0^b \cos^{-1}\left[\frac{d-\lambda d+\sqrt{d^2+x^2}}{\lambda d}\right]dx, \\ &= \frac{bp_0}{2}\sqrt{\frac{\lambda-1}{\lambda}}, \end{aligned} \tag{29}$$

the evaluation being readily effected by an integration by parts, followed by two elementary and obvious substitutions of variable.

The moment of the uplift forces, relative to the heel of the dam, is

$$M = \frac{p_0}{\pi}\int_0^b x\cos^{-1}\left[\frac{d-\lambda d+\sqrt{d^2+x^2}}{\lambda d}\right]dx.$$

This expression can also be evaluated by integration by parts, followed by several standard substitutions. The result is

$$(30)\quad M = \frac{b^2 p_0}{16\pi}\left[\frac{2(3\lambda - 2)}{\lambda\sqrt{\lambda - 1}} + \frac{3\lambda - 4}{\lambda - 1}\left(\pi - 2\sin^{-1}\frac{1}{\sqrt{\lambda}}\right)\right].$$

The point of application of the resultant uplift force is thus located, from the heel, a distance:

$$(31)\quad X = \frac{b}{8\pi(\lambda - 1)} \times\left[\frac{2(3\lambda - 2)}{\sqrt{\lambda}} + (3\lambda - 4)\sqrt{\frac{\lambda}{\lambda - 1}}\left(\pi - 2\sin^{-1}\frac{1}{\sqrt{\lambda}}\right)\right].$$

All of the formulas of this section may readily be modified to provide for a non-vanishing tail-water pressure. In (25), for example, one must replace p_0 by $p_0 - p_1$; and must add p_1. The curves of Fig. 5, for example, may then be interpreted as giving the per cent of the difference between head and tail-water pressure which must, at any point, be added to the tail-water pressure to give the total pressure.

From equation (29) one may compute the total uplift force (per foot of dam) for various depths of sheet-piling at the heel. These results are shown in Fig. 6. The ordinates plotted repre-

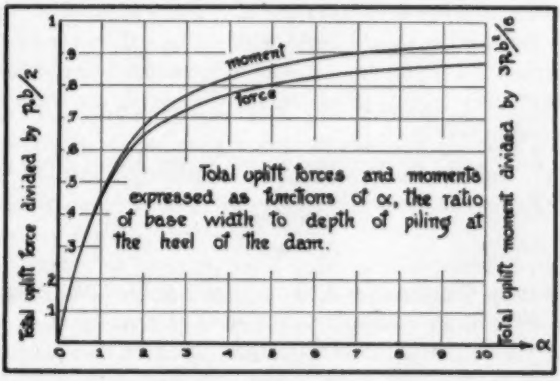


FIG. 6

sent the total uplift forces divided by $p_0 b/2$; that is to say, the ordinates represent fractions of the "normal" uplift pressures which would obtain were there no sheet-piling. Figure 6 also shows the dependence of the uplift moment on the ratio α of base width to piling depth. In the case of the moment curve, the ordinates plotted are actual moments divided by $3p_0 b^2/16$. Thus (see equation 14) the ordinates measure fractions of the normal uplift moments which would obtain were no sheet-piling present.

§ 7. Discussion: Line of Creep Theory

The above formulas permit one accurately to discuss the errors involved in the assumption of linear distribution of pressures and the "line of creep" theory. As regards pressures, the line of creep theory assigns, in the case of piling at the heel, a pressure which drops linearly to zero from the initial value:

$$\frac{pb_0}{2d+b} = \frac{\alpha}{2+\alpha} p_0.$$

This initial value is uniformly larger than that given by the accurate formula (12). For $\alpha = 1$ the error is $+22.4$ per cent; for $\alpha = 2$, $+18$ per cent; while for higher values of α the error fluctuates slowly about $+16$ per cent. It is clear from Fig. 5, however, that a linear distribution of pressure approximating the actual distribution should begin with initial values considerably higher than the actual ones. A more substantial test of the line of creep theory is thus to be found by studying the total uplift forces and moments.

The total uplift force corresponding to the line of creep distribution is

$$\frac{\alpha}{2+\alpha} \frac{bp_0}{2}.$$

This is to be compared with the accurate value (29). Figure 7 shows, for various values of α (the ratio of dam width to piling depth), the per cent of the true value (29) which results from the approximate "line of creep" expression for total uplift force.

It should be remembered that the present discussion concerns only the case of piling at the heel of the dam. It is noted from Fig. 7 that the line of creep theory assigns total uplift forces which are but 80 per cent to 90 per cent of the actual total uplift

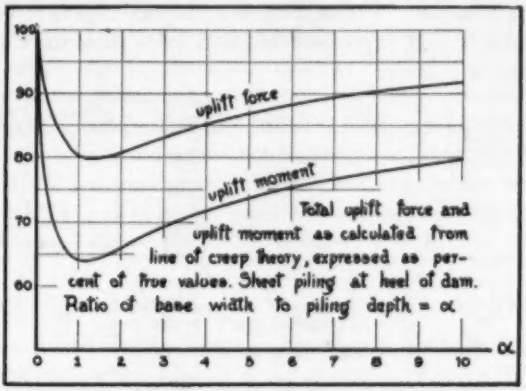

100
90
80
70
60
uplift force
uplift moment
Total uplift force and uplift moment as calculated from line of creep theory, expressed as percent of true values. Sheet piling at heel of dam. Ratio of base width to piling depth = α
0 1 2 3 4 5 6 7 8 9 10
α

FIG. 7

force, so that the assumption is unfortunate from a practical point of view.

The moment assigned by the line of creep theory is

$$\frac{\alpha}{\alpha+2}\frac{p_0 b^2}{6},$$

which is to be compared with the accurate expression (31). Figure 7 also shows the per cent of the true moment which is given by the approximate expression just written. The approximate moments are, for ordinary α, from 65 per cent to 80 per cent of the true moments.

The ranges of error in total uplift and moment depend on the depth of piling as compared with the width of the base. If piling were driven deeper than the base width ($\alpha < 1$) the errors in the line of creep assumption would be larger than those quoted above.

§ 8. Discussion: Lines of Seepage Flow

Although the pressure on the base of the dam is the immediate consideration, it is of considerable interest and importance to study, by means of equations (23) the lines of flow and the lines of equal pressure. Although these equations appear to be somewhat complicated, it proves feasible to solve them for x and y as functions of u and v. In fact,

$$x = \pm\left[\frac{U^2 - V^2 - d^2 + \sqrt{(U^2 - V^2 - d^2)^2 + 4U^2V^2}}{2}\right]^{1/2}, \quad \begin{matrix} + \text{ for } U > 0 \\ - \text{ for } U < 0 \end{matrix} \tag{32}$$

$$y = +\left[\frac{d^2 + V^2 - U^2 + \sqrt{(U^2 - V^2 - d^2)^2 + 4U^2V^2}}{2}\right]^{1/2},$$

where

$$U = d\lambda \cosh u \cos v + d\lambda - d,$$
$$V = d\lambda \sinh u \sin v.$$

A graph of the curves u = constant and v = constant is shown, in Fig. 8, for the case of a dam whose base width is twice the piling

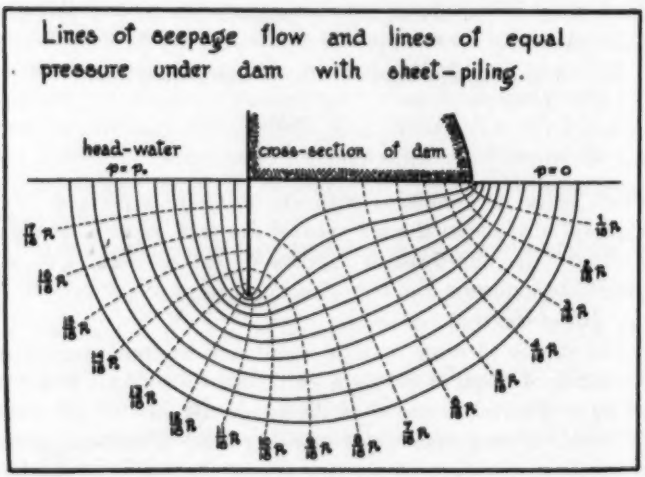


Fig. 8

depth. It is interesting to note that the distortion which the sheet-piling produces in the lines of flow does not extend as far from the piling as one might expect. In fact, the line of flow farthest from the dam, in the figure, does not differ far from the ellipsoidal curve which would obtain if the piling were absent. The distortion due to the piling extends a distance in all directions equal, roughly, to the piling depth. The interest in this observation lies in the strong suggestion that non-homogeneities in foundation material very likely are effective in disturbing seepage flow only in a neighborhood whose dimensions are roughly twice those of the non-homogeneity. One thus has a crude means for judging when and to what extent is applicable an analysis based on homogeneity.

§ 9. Discussion: Sheet-Piling Not at the Heel

If the sheet-piling be located a distance b_1 from the heel of the dam, the parameters λ and $\varkappa$ of the general expression (17) are to be calculated from (19) and (20). Thus, setting

$$\alpha_1 = \frac{b_1}{d}, \tag{33}$$

one has

$$\begin{aligned} \varkappa &= \sqrt{1 + \alpha_1^2}, \\ \lambda &= \frac{\sqrt{1 + \alpha_1^2} + \sqrt{1 + (\alpha - \alpha_1)^2}}{2}. \end{aligned} \tag{34}$$

Taking the real part of equation (17), and setting $u = 0, y = 0$ in order to find the pressure on the base of the dam,

$$d(\lambda \cos v + \lambda - \varkappa) = \pm \sqrt{x^2 + d^2},$$

or

$$p(x, y) = \frac{p_0}{\pi} c(x, y) = \frac{p_0}{\pi} \cos^{-1}\left[\frac{d(\varkappa - \lambda) \pm \sqrt{x^2 + d^2}}{d\lambda}\right], \tag{35}$$

the plus sign holding for $x \geq 0$, the minus sign for $x \leq 0$, and the origin $x = 0$ corresponding to the position of the piling.

Figure 9 shows the distribution of pressure across the base of a dam whose sheet piling is as deep as the dam is wide ($\alpha = 1$). The various curves correspond to various locations of the sheet-piling, the parameter α_1 being given the values 0, 0.1, 0.2, $\cdots$ 1.0.

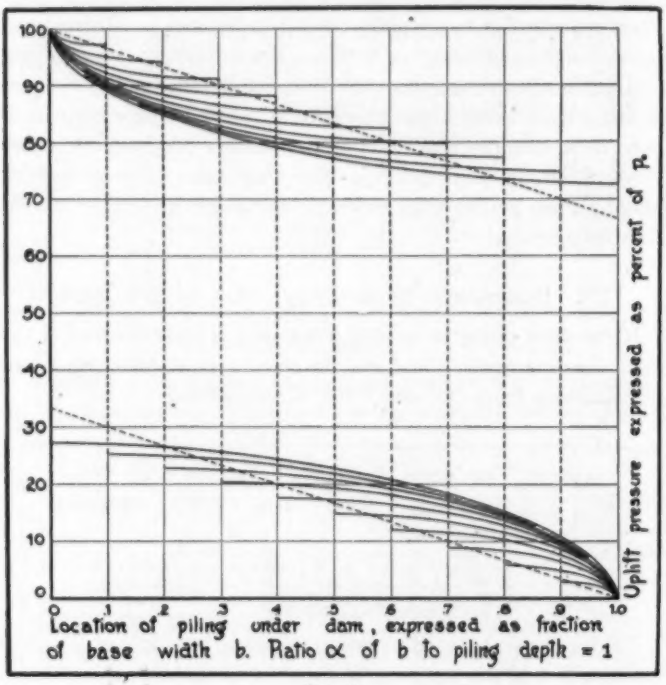


FIG. 9

No matter where the sheet-piling is relative to the base of the dam, the origin $x = 0$ of equation (35) is located at the sheet-piling. In the figures, however, it is more natural to hold the dam fixed and move the piling. The position of the piling, corresponding to any curve, is indicated by the vertical dotted portion of the curve. The length of this dotted portion gives directly

the drop of pressure across the piling. Figures 10, 11, and 12 show similar curves for other depths of piling ($\alpha = 2, 4, 8$, respectively).

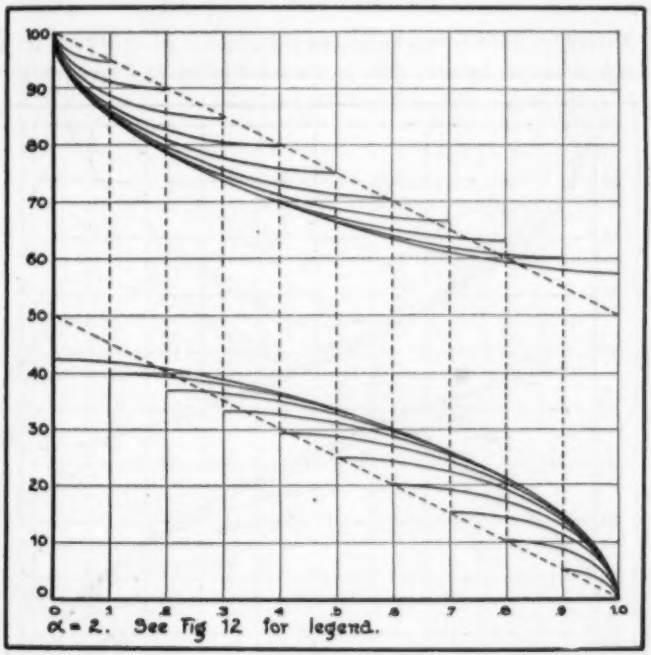


FIG. 10

Consider cases in which α has the same value, but in which the piling is differently located. In the first case, let the piling be b_1 feet from the heel, and in the second case, $b_2' = b_1$ feet from the toe. All quantities which refer to the second case will be represented by primed letters. One has

$$\alpha' = \alpha - \alpha_1,$$

so that

$$\lambda' = \lambda$$

$$\varkappa' - \lambda' = \frac{\sqrt{1 + \alpha_1^2} - \sqrt{1 + (\alpha - \alpha_1')^2}}{2}$$

$$= \frac{\sqrt{1 + (\alpha - \alpha_1)^2} - \sqrt{1 + \alpha_1^2}}{2}$$

$$= -(\varkappa - \lambda).$$

Consider, moreover, the pressure $p(\xi)$ at a point which is, in the first case a distance ξ from the sheet-piling (ξ either positive

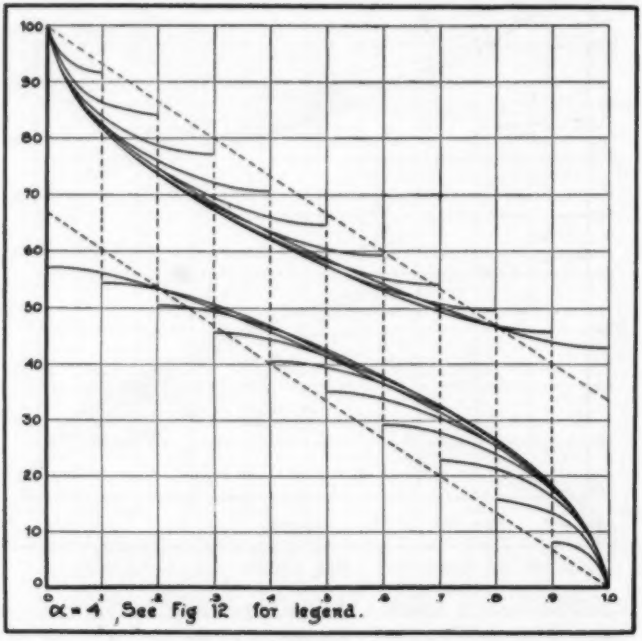


FIG. 11

or negative) so that $x = \xi$; and the pressure $p'(-\xi)$ at a point which is in the second case a distance $-\xi$ from the piling so that $x' = -\xi$. Then

$$p(\xi) = \frac{p_0}{\pi} \cos^{-1}\left[\frac{d(\varkappa - \lambda) \pm \sqrt{2 + d^2}}{\lambda d}\right],$$

$$p'(-\xi) = \frac{p_0}{\pi}\cos^{-1}\left[\frac{-d(\varkappa-\lambda) \mp \sqrt{2+d^2}}{\lambda d}\right],$$

the sign of the radical being opposite in the two equations. Thus

(36) $$p(\xi) = p_0 - p'(-\xi).$$

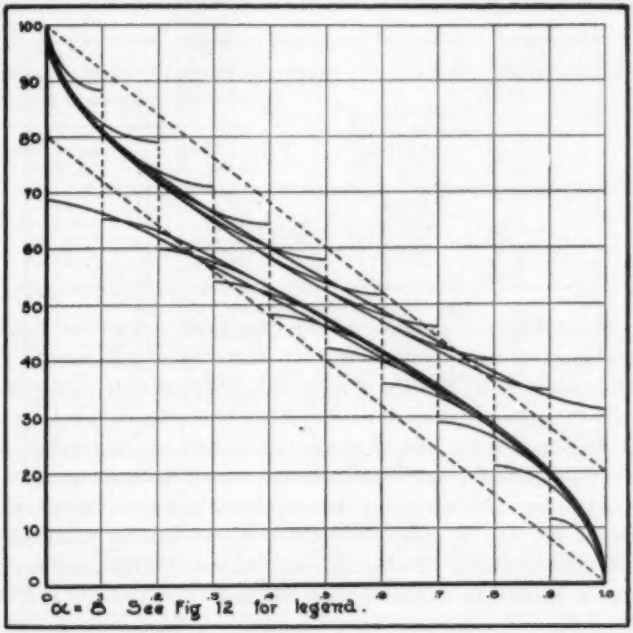


FIG. 12

By virtue of this relationship it was only necessary, in obtaining the data for Figs. 9 to 12, to compute for values of α_1 from zero to one-half.

The drop of pressure at the sheet-piling is, on the basis of the line of creep assumption,

$$\frac{2}{2+\alpha}p_0,$$

and is independent of the location of the piling. The drop, as calculated from (35), varies with b_1 as well as with α. In fact, the drop is larger, the farther the piling is from the center of the base, the drop being the same in two cases in which the piling is the same distance from, but on opposite sides of, the center. The range of error is indicated in the following table:

TABLE II

α	"Line of creep" drop of pressure at piling	Actual drop of pressure at piling	
		Piling at heel or toe	Piling at center of base
1	.667 p_0	.728 p_0	.704 p_0
2	.500 p_0	.576 p_0	.500 p_0
4	.333 p_0	.430 p_0	.296 p_0
8	.200 p_0	.311 p_0	.156 p_0

These data are presented very briefly since they are not of fundamental significance. In fact, the *drop of pressure* at the piling assigned by the line of creep theory differs only moderately from the accurate drop in pressure; but this remark is somewhat misleading, for the line of creep theory assigns, in general, too small a value to the pressure immediately on the head-water side of the piling. As before, one makes a fairer and more illuminating test of the line of creep assumption by comparing total uplift force and moment. The general expression (35) for the pressure cannot readily be integrated, so the total uplift forces and the moments were obtained, from the curves of Figs. 9 to 12, partly by mechanical and partly by computational quadrature. Figure 13 shows the total uplift force, expressed in percentage of $bp_0/2$, the normal value when sheet-piling is not present. The abscissas locate the position of the sheet-piling under the base of the dam. It is noted that when the piling is under the center of the dam, the total uplift force has precisely the normal value it would have if the piling were not present. That is, the piling shields the down-stream half of the base from the head-water pressure to

just the same extent that it shields the upstream half of the base from the (zero) tail-water pressure. When the piling is located under the down-stream half of the base, the total uplift forces are greater than the normal value, since then the piling is being effectively used to shield the base of the dam from the zero tail-water pressure, rather than to shield it from the head-water

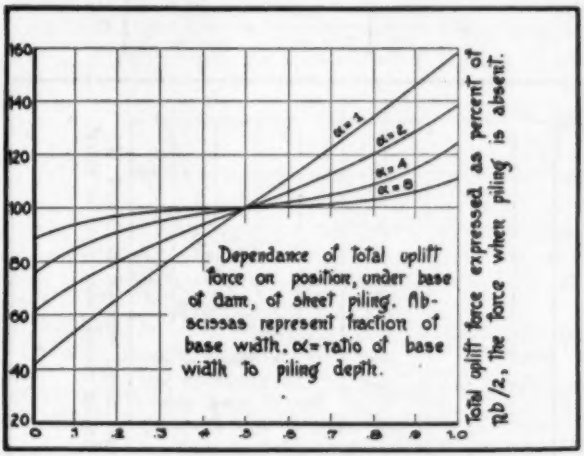


FIG. 13

pressure. The various curves of Fig. 13 are drawn for different ratios α of base width to piling depth.

Similarly, Fig. 14 shows the dependence of the total moment (relative to the heel of the dam) of the uplift force upon the position of the piling under the base.

The total uplift force and moment (relative to the heel of the dam) which result from the line of creep assumption are readily calculated by integration. Suppose the piling is at a distance b_1 from the heel, and set

$$\beta = \frac{b_1}{b}.$$

The parameter β is precisely that used as an abscissa in Figs. 9 to 12. The total "line of creep" uplift force is then

$$F = \frac{pb_0}{2}\left[\frac{\alpha^2}{\alpha(\alpha+2)} + \frac{4\beta}{\alpha+2}\right], \tag{37}$$

while the "line of creep" moment is

$$M = \frac{3b^2p_0}{16}\left[\frac{8\alpha}{9(\alpha+2)} + \frac{16\beta^2}{3(\alpha+2)}\right]. \tag{38}$$

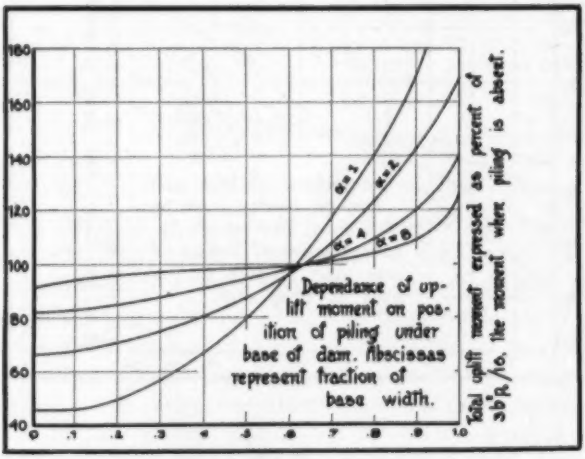


FIG. 14

From these expressions one can readily calculate the "line of creep" forces and moments for any ratio α of base width to piling depth, and for any position β of the piling under the dam. These calculated values may then be compared with the accurate values plotted in Figs. 13 and 14, and the errors in the line of creep assumption may be determined.

The results of this calculation will not be given in detail, since it is possible to summarize these results briefly and sufficiently accurately. As regards total uplift force, the line of creep theory

gives results too small for $\beta < .5$, and too large for $\beta > .5$. For $\beta = 0$ (piling at the heel) the error ranges from 20 per cent for $\alpha = 1$ to 10 per cent for $\alpha = 8$. At $\beta = .5$ the error is zero for any value of α. For $\beta = 1.0$, the error is about 8 per cent, and does not depend very sensitively on the value of α. As regards moments, the line of creep assumption gives results too low for $\beta < .7$, and very slightly too high (of the order of 5 per cent or less) for $\beta > .7$. For $\beta = 0$ the error is about 30 per cent for all values of α, this error reducing approximately linearly to zero at $\beta = .7$.

THE RANK OF A MATRIX, THE VALUE OF A DETERMINANT, AND THE SOLUTION OF A SYSTEM OF LINEAR EQUATIONS

By Lepine Hall Rice

Starting with an $m \times n$ matrix $A = ||a_{ij}||$ of rank r, let us first perform certain elementary transformations, which will not change the rank of A.

Choosing any nonzero element a_{pq}, we multiply by it every row excepting the pth row. We next subtract from the ith row ($i = 1, \cdots, p-1, p+1, \cdots, m$) the pth row multiplied by a_{iq}. This will leave the pth row unchanged, will reduce every element of the qth column to zero with the exception of a_{pq}, and will replace every other element a_{ij} of A by $a_{ij}a_{pq} - a_{pj}a_{iq}$. Finally we multiply by -1 each row beyond the pth row and each column beyond the qth column. The expression replacing each a_{ij} is now identical with the 2-rowed determinant of A containing a_{ij} and a_{pq}, and the new matrix A', with a'_i written for a_{ij}, is:

$$\left\|\begin{array}{cccccccccccc}
\begin{vmatrix} a_1^1 & a_1^q \\ a_p^1 & a_p^q \end{vmatrix} & \cdots & \begin{vmatrix} a_1^{q-1} & a_1^q \\ a_p^{q-1} & a_p^q \end{vmatrix} & \begin{vmatrix} a_1^q & a_1^{q+1} \\ a_p^q & a_p^{q+1} \end{vmatrix} & \cdots & \begin{vmatrix} a_1^q & a_1^n \\ a_p^q & a_p^n \end{vmatrix} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\begin{vmatrix} a_{p-1}^1 & a_{p-1}^q \\ a_p^1 & a_p^q \end{vmatrix} & \cdots & \begin{vmatrix} a_{p-1}^{q-1} & a_{p-1}^q \\ a_p^{q-1} & a_p^q \end{vmatrix} & \begin{vmatrix} a_{p-1}^q & a_{p-1}^{q+1} \\ a_p^q & a_p^{q+1} \end{vmatrix} & \cdots & \begin{vmatrix} a_{p-1}^q & a_{p-1}^n \\ a_p^q & a_p^n \end{vmatrix} \\
a_p^1 & \cdots & a_p^{q-1} & a_p^q & -a_p^{q+1} & \cdots & -a_p^n \\
\begin{vmatrix} a_p^1 & a_p^q \\ a_{p+1}^1 & a_{p+1}^q \end{vmatrix} & \cdots & \begin{vmatrix} a_p^{q-1} & a_p^q \\ a_{p+1}^{q-1} & a_{p+1}^q \end{vmatrix} & \begin{vmatrix} a_p^q & a_p^{q+1} \\ a_{p+1}^q & a_{p+1}^{q+1} \end{vmatrix} & \cdots & \begin{vmatrix} a_p^q & a_p^n \\ a_{p+1}^q & a_{p+1}^n \end{vmatrix} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\begin{vmatrix} a_p^1 & a_p^q \\ a_m^1 & a_m^q \end{vmatrix} & \cdots & \begin{vmatrix} a_p^{q-1} & a_p^q \\ a_m^{q-1} & a_m^q \end{vmatrix} & \begin{vmatrix} a_p^q & a_p^{q+1} \\ a_m^q & a_m^{q+1} \end{vmatrix} & \cdots & \begin{vmatrix} a_p^q & a_p^n \\ a_m^q & a_m^n \end{vmatrix}
\end{array}\right\|.$$

Now form an $(m-1) \times (n-1)$ matrix A_1 by striking out the pth row and qth column of A'.

The rank of A_1 is $r-1$.

For, corresponding to any nonvanishing k-rowed determinant K of A' not containing elements of the pth row, there is a non-

vanishing $(k+1)$-rowed determinant of A' of which K is a first minor and a_{pq} is its complement. Hence any nonvanishing r-rowed determinant of A' must contain elements of the pth row; and must contain at least one $(r-1)$-rowed nonvanishing minor which is a determinant of A_1.

It has long been known that a determinant can be condensed to the next lower order by the formula

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdot & \cdot \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \frac{1}{a_{11}^{n-2}} \begin{vmatrix} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} & \cdots & \begin{vmatrix} a_{11} & a_{1n} \\ a_{21} & a_{2n} \end{vmatrix} \\ \cdot & \cdot & \cdot \\ \begin{vmatrix} a_{11} & a_{12} \\ a_{n1} & a_{n2} \end{vmatrix} & \cdots & \begin{vmatrix} a_{11} & a_{1n} \\ a_{n1} & a_{nn} \end{vmatrix} \end{vmatrix}.$$

But from what precedes it is clear that any convenient nonzero element a_{pq} may be chosen as the base of condensation, since $(-1)^{p+q}$ is exactly used in changing the signs of the rows and columns beyond a_{pq}. That is, when A is an $n \times n$ matrix,

$$|A| = \frac{1}{a_{pq}^{n-2}} |A_1|.$$

To show how the foregoing may be applied to the solution of a system of linear equations, we give a simple illustration. Let the system be

$$\begin{aligned} 5x - 2y - 2z - 2w &= 3, \\ 2x + 5y + 7z + 3w &= 12, \\ 2x + 3y + 4z + w &= 7, \\ 7x - y - z - 3w &= 5, \end{aligned}$$

The augmented matrix is

$$\begin{matrix} 5 & -2 & -2 & -2 & 3 \\ 2 & 5 & 7 & 3 & 12 \\ 2 & 3 & 4 & 1\rceil_1 & 7 \\ 7 & -1 & -1 & -3 & 5 \end{matrix} \qquad (1)$$

The mark "$\rceil_1$" denotes that the element 1 is chosen as the first base of condensation. Replacing the elements of the remaining

rows and columns by the 2-rowed determinants containing the base, it is well to keep everything in position and carry along the base for reference, thus:

$$\begin{array}{rrrrr} 9 & 4\rceil_2 & 6 & \cdot & -17 \\ -4 & -4 & -5 & \cdot & 9 \\ \cdot & \cdot & \cdot & 1\rceil_1 & \cdot \\ -13 & -8 & -11 & \cdot & 26 \end{array} \tag{2}$$

We now make the second condensation with the base 4 marked above.

$$\begin{array}{rrrrr} \cdot & 4\rceil_2 & \cdot & \cdot & \cdot \\ -20 & \cdot & 4 & \cdot & -32 \\ \cdot & \cdot & \cdot & 1\rceil_1 & \cdot \\ -20 & \cdot & 4 & \cdot & -32 \end{array} \tag{3}$$

As (3) is of rank 1 both with and without the last column, the common rank of the matrix of the coefficients and the augmented matrix is 3. Had the last column of (3) not been proportional to the others, the equations would have been inconsistent.

As the determinant of the coefficients of y, z, w in the first three equations is seen from (3) to have the value $4/(1^1 4^0) = 4$, we proceed to solve these equations for y, z, and w as linear functions of x.

Finding first the coefficients of x in the functions, that for $4y$ is $-\begin{vmatrix} 9 & 6 \\ -4 & -5 \end{vmatrix} = 21$, in (2). That for $4z$ is -20, in (3). That for $4w$ is found to be 9.

The absolute term for $4y$ is $-\begin{vmatrix} 6 & -17 \\ -5 & 9 \end{vmatrix} = -31$, in (2). That for $4z$ is $-(-32) = 32$, in (3). That for $4w$ is found to be -7.

Thus the complete solution is $4y = 21x - 31$, $z = -5x + 8$, $4w = 9x - 7$.

What follows has no necessary relation to equations. When in A_1 a nonzero element is chosen as a second base and the process of condensation is repeated (the scheme of arrangement

in (1)–(2)–(3) being used), giving an $(m-2)\times(n-2)$ matrix A_2, and again the process is worked on A_2, and so on, we reach an $(m-r+1)\times(n-r+1)$ matrix A_{r-1} of rank 1. If the successive bases have the values $b, b_1, \cdots, b_{r-2}$, then any nonzero element b_{r-1} of A_{r-1}, divided by $b^{r-2}{b_1}^{r-3}\cdots b_{r-3}$, is the value of that r-rowed determinant of A whose rows and columns are those (in our scheme of arrangement) of $b, b_1, \cdots, b_{r-3}, b_{r-2}, b_{r-1}$.

That is, the method not only determines the rank r of the matrix but picks out and evaluates a nonvanishing r-rowed determinant of the matrix. It also facilitates the evaluation of a number of other determinants of the matrix.

Whatever nonzero element of A be chosen as b, we can find a b_1, and so on to b_{r-1}. Thus we have here a proof that any nonzero element of a matrix of rank r is an element of at least one nonvanishing r-rowed determinant of the matrix.

A RECENT CONTRIBUTION TO THE DESIGN OF ELECTRIC FILTER NETWORKS

By E. A. Guillemin

1. *Introductory Remarks.* Early last year a new method for the design of electric wave filters was published by W. Cauer,[1] a German engineer who is well known in the field of network synthesis for various notable contributions. Although this new method has several points of advantage over that developed principally by O. J. Zobel [2] for the design of filters of the Campbell type, it has thus far received little attention or has remained wholly unknown to a good many American engineers who are working in this field. The reason is probably due to the form of Cauer's publication, which is quite mathematical and in general not in accord with the manner in which similar material is presented in this country. It is, therefore, rather difficult to follow by the lay engineer to whom the method might otherwise prove useful.

The object of the present paper is to outline the Cauer method of design in a manner which is not as rigorous and general as that used in Cauer's own publication, but on the other hand can be more readily understood and applied with sufficient accuracy for most practical cases. At appropriate points it shall also be our object to show the relation of the results obtainable by this new method to those given by previous design methods and their subsequent development, referred to above. It is assumed, of

[1] *Siebschaltungen*, published by V. D. I., Verlag G. M. B. H. Berlin, 1931. Since writing this paper, a condensed version of this theory has appeared in *Physics*, vol. 2, No. 4, April 1932, p. 242.

[2] O. J. Zobel: "Theory and Design of Uniform and Composite Electric Wave Filters," *B. S. T. J.*, Jan. 1923; "Transmission Characteristics of Electric Wave-filters," *B. S. T. J.*, Oct. 1924; "Extensions to the Theory and Design of Electric Wave Filters," *B. S. T. J.*, April 1931.

course, that the reader is sufficiently familiar with the general aspects of this problem so that detailed discussion of matters in common with the subject at large may be omitted. For these the reader is referred to the second reference above as well as to several texts [3] which deal with this subject.

In the design of electric filter networks, two major considerations receive our attention. The first, and perhaps most important, is that the structure shall meet certain minimum requirements regarding attenuation as well as definite requirements regarding cut-off points and the location of pass bands in the frequency spectrum. The second is that the structure shall present, at its input and output terminals, a characteristic impedance which is as nearly equal to a constant nominal value over the essential frequency range as it is possible to obtain with a given amount of material and expense. Indeed, the premises upon which the theory of operation are based, make this second requirement a necessary condition for the exact fulfillment of the first. In general, however, it is found that the predicted attenuation properties will be met with sufficient accuracy if the characteristic impedance is substantially equal to a constant nominal value over most of the pass band of the filter. A specific problem makes it necessary to be able to specify definite minimum requirements as to the frequency limits within which the characteristic impedance shall approximate a constant nominal value with a prescribed degree of tolerance. Thus we see that the two outstanding requirements of a design method are that it shall enable us to prescribe independently certain limits and tolerances regarding the *attenuation* and *characteristic impedance* functions.

In the case of the uniform ladder structure, which is basic for the Campbell type of filter, the analytic expressions for the characteristic impedance and attenuation functions are unfortunately of such a nature that a design method which attempts to

[3] See, for example, K. S. Johnson, *Transmission Circuits for Telephonic Communication* and T. E. Shea, *Transmission Networks and Wave Filters*, D. Van Nostrand, 1929.

go out from prescribed behaviour on the part of these functions becomes exceedingly involved if not altogether impossible. For example, for a T-section we have the well known expressions for characteristic impedance and propagation functions respectively:

$$Z_0 = \sqrt{Z_1 Z_2}\sqrt{1 + \frac{Z_1}{4Z_2}} \tag{1}$$

and

$$\gamma = 2ln\left(\sqrt{\frac{Z_1}{4Z_2}} + \sqrt{1 + \frac{Z_1}{4Z_2}}\right), \tag{2}$$

where Z_1 and Z_2 are the series and shunt impedances of the section. Since

$$\gamma = \gamma_1 + j\gamma_2 \tag{3}$$

we have for the attenuation function:

$$\gamma_1 = 2ln \Big/ \sqrt{\frac{Z_1}{4Z_2}} + \sqrt{1 + \frac{Z_1}{4Z_2}} \Big/ . \tag{2a}$$

The logical design process would be to determine what form of functions may be chosen so as to eventually lead to physically realizable impedances Z_1 and Z_2 in conformance with the expressions (1) and (2), and then attempt to adjust the various parameters contained in these functions so as to meet the behaviour required by the problem under consideration. Although the functions (1) and (2) are not so complicated as to make this process seem impossible, they are nevertheless sufficiently complex to discourage anyone from attempting to base a design method upon such a procedure.

Thus the method of designing filters based upon the ladder structure has taken quite a different aspect from the one just described. In fact the question of minimum tolerances is for the moment disregarded altogether in an attempt to arrive at a structure which will at least be a first approximation to the desired result. This first approximation is known as the "constant-k" type, for which the impedances Z_1 and Z_2 are reciprocal

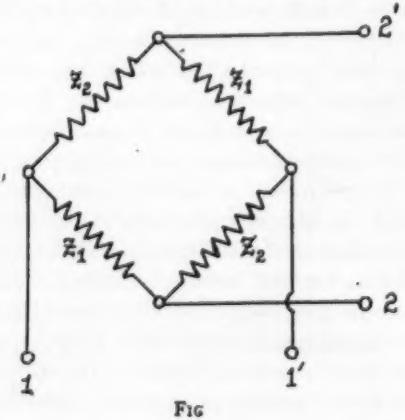
in nature. The structure for this type has the advantages of being simple in form, and the parameters in it are easy to compute. The chief disadvantages are that the attenuation function does not rise rapidly enough beyond cut-off, and that the characteristic impedance function is not nearly constant enough throughout the pass band for many practical cases.

The succeeding steps in the design process consist in determining additional sections to be connected in series with this first-approximation section so as to improve either, or perhaps both of the shortcomings of the latter. These additional sections, which are called M-derived types, are so derived that they will operate in conjunction with the first section without introducing any impedance irregularities. However, it appears that the first additional section can be designed with a view toward improving either the resulting characteristic impedance function or the attenuation function, but not both, except in the nature of a compromise. If this compromise is insufficient for the problem in hand, then further derived sections must be added, and so forth until all requirements are met.

The method therefore is essentially a patch-work process which eventually satisfies all requirements. This in itself is not sufficient reason to belittle the method, since many respectable design processes in engineering work involve just such a line of attack. The chief dissatisfaction which the student finds in acquainting himself with this method lies in the seemingly illogical or hit and miss approach. Pedagogically it is difficult to present such a method because it lacks a logical trend. Some results seem to be arrived at as if by accident. For example, in the design of the first derived section we may say that we are attempting to improve the sharpness of the cut-off properties, in which the "constant-k" section is lacking. When we afterward find that the additional section also possesses an improved characteristic impedance function when properly split and connected to the first section, we wonder whether this is merely good fortune or whether the method was originally intended to produce this result.

The fundamental difficulty with this method obviously lies in the fact that it does not enable us in any one of the steps to specify the attenuation and characteristic impedance functions *independently* of each other. The two are always entangled, and by some happy chance in such a way that an improvement in one carries a partial improvement in the other with it. Added to this is of course the question as to whether the final result obtained is the most economical one regarding the total number of coils and condensers. Our intuition would seem to tell us that a process which hits and misses as it proceeds will very likely be less economical in the end than one which proceeds logically from the given requirements to a single structure which meets all of these. The Cauer method of design, which we shall now take up, does proceed logically, and in some cases leads to better economy.

2. *General Aspects of the Method.* The fact that the characteristic impedance and attenuation functions are so intimately related in the Campbell type of filter, is due to the ladder structure upon which they are based. Cauer obviates this inherent difficulty by choosing the symmetrical lattice or bridge structure as basic. This is illustrated in Fig. 1. 1 and 1′ are input



FIG

terminals, and 2 and 2′ are the output terminals. It is well known that for this structure the characteristic impedance and attenuation functions respectively are given by

$$Z_0 = \sqrt{Z_1 Z_2} \tag{4}$$

and

$$\gamma = ln\left(\frac{\sqrt{\frac{Z_1}{Z_2}} + 1}{\sqrt{\frac{Z_1}{Z_2}} - 1}\right). \tag{5}$$

Here we see that the characteristic impedance function is determined by the product of the two component impedances, while the attenuation function is determined by their quotient. Thus it is obvious that, having specified these two fundamental functions independently, we obtain the analytic expressions for the component impedances by simply forming their product and quotient. These functions must, of course, be selected in such a form that the resulting functions for the impedances Z_1 and Z_2 are physically realizable. This may easily be accomplished by starting with a series of analytical expressions which we know represent physical impedance structures, and using these to generate a corresponding series of the functions

$$\sqrt{Z_1 Z_2}$$

and

$$\sqrt{\frac{Z_1}{Z_2}}.$$

Assuming then that this series has been completely determined, our design process resolves itself into the selection of any two functions out of this series, and the determination of their various parameters in accordance with the desired behaviour as required by the particular problem in hand. Once this has been done, the functions Z_1 and Z_2 are easily found by multiplication and division respectively, and their physical realization can then be carried out by means of any one of the several methods which

have been developed for this purpose.[4] This general procedure we shall now formulate more definitely.

To begin with, we wish to state that we shall not bother with taking the effect of incidental resistances into account in this paper. For this process the reader is referred to Cauer's own publications or to his own resourcefulness, for this is a matter which can be taken care of by the judicious application of the usual formulae which apply to lattice structures, and has nothing to do with the fundamentals of the design method itself. We shall concern ourselves only with these, and therefore consider only pure reactance networks in all that follows.

Let us start by elaborating somewhat upon the expressions (4) and (5). The first is the characteristic impedance function. We want it to be real in the pass band. It should also be as nearly constant as possible of course, but to begin with it should at least be real. This means that, since Z_1 and Z_2 are pure imaginaries, they should have opposite signs in a pass band. Hence we recognize one requisite for the impedances Z_1 and Z_2 already, namely, that they should be reciprocal in nature in any pass band or transmission region, because reciprocal impedances always have opposite signs. How does this check with regard to the attenuation function (5)? Here we see that if Z_1 and Z_2 are of opposite sign, the function

$$\sqrt{\frac{Z_1}{Z_2}}$$

becomes a pure imaginary. Consequently the numerator and denominator of the argument of the logarithm are conjugate complex numbers, and the magnitude of their ratio is, therefore, unity. The real part of the logarithm is zero. Hence the attenuation is zero as it should be for a transmission region. This is fortunate.

[4] R. M. Foster, "A Reactance Theorem," *B. S. T. J.*, April 1924. W. Cauer, "Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit," *Archiv f. Elektrotechnik*, vol. 17, p. 355, 1927.

Now for the attenuation region. Here we wish to have the attenuation as high as possible. Examining (5) we see that this expression will become infinite if the component impedances become equal. If they cannot be equal they should at least have the same sign, because, as soon as they do, the attenuation becomes greater than zero. Hence an attenuation region occurs wherever Z_1 and Z_2 are alike in sign and tend to be more or less equal. The characteristic impedance becomes purely imaginary under these conditions and hence cannot match the same resistance at the terminals of the structure that it was intended to match over the pass band. However, if the attenuation is very high, the mismatching situation becomes less serious. At all events it is much more important to match properly in a transmission region.

To summarize, we see that wherever Z_1 and Z_2 are reciprocal, the characteristic impedance is real and equal to unity, while the attenuation is zero; and wherever Z_1 and Z_2 are equal the attenuation is infinite, while the characteristic impedance is purely imaginary and may have any value, but on account of the infinite attenuation we do not let this disturb us particularly. Hence in a pass band we should see to it that Z_1 and Z_2 are as nearly reciprocal as possible, and in an attenuation region we should see to it that these impedances are as nearly equal as possible. A cut-off point will therefore form the dividing line between these two tendencies.

It may be appreciated in a general way that it would be too much to expect two physical impedances to be exactly reciprocal over a certain region of frequencies, and to switch over suddenly at some point so as to be exactly equal in the region beyond. We rather expect that the impedances will be neither exactly reciprocal in the one region nor exactly equal in the other. Also we expect that the transition from the one tendency to the other will be gradual, but that the tendencies will become more complete farther from the transition point. Furthermore, we can perhaps also predict that the transition will become more

abrupt, the more poles and zeros the impedance functions have in the vicinity of the transition point. Just how the transition takes place, and how nearly the impedances become reciprocal on the one hand or equal on the other, is a matter which will require some study of these functions. In any case, our problem is to determine the necessary poles and zeros of the component impedances as well as their most advantageous distribution so that they may meet the requirements stipulated for that case.

Let us illustrate this with a simple example. Before we start, however, we wish to make the following remarks relative to the manner in which we shall indicate graphically the general character of an impedance function. It is well known that a physically realizable dissipationless impedance function is regular and continuous except for a finite number of simple poles, the latter being separated by simple zeros. Furthermore, the slope of the impedance function is everywhere positive. Such an impedance function is uniquely determined by the specification of the location of its poles and zeros plus one additional piece of information which may take the form of a multiplying factor. Since, therefore, most of the characteristic behavior of the function is determined by the allocation of its poles and zeros, it is sufficient for our present purpose to merely indicate in a figure the location of poles and zeros on a frequency axis, and omit drawing in the function itself. This scheme saves time and space, and at the same time leads to an easier understanding of the principles which we wish to illustrate. In this connection we shall adopt the convention of denoting the location of a zero by a circle and that of a pole by a cross.

Thus, for example, Fig. 2 represents two impedances Z_1 and Z_2, of which the first is obviously a coil and condenser in series, and the second simply a coil. The zero of Z_1 is numbered 1. The angular frequency is denoted by ω, and the point $\omega = \infty$ is indicated as a finite point for convenience. We see that Z_1 and Z_2 have opposite signs for $0 < \omega < \omega_1$,* and that they have

* The subscript on ω refers it to the correspondingly numbered critical frequency in the figure.

like signs for $\omega_1 < \omega < \infty$. Hence, if we use these impedances in a lattice structure, the latter becomes a low-pass filter with cut-off at ω_1. In order to be able to tell how the characteristic impedance varies over the pass band, and how the attenuation

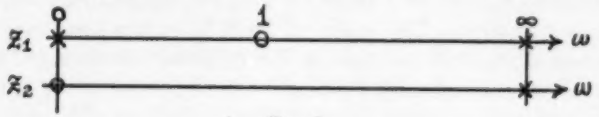


FIG. 2.

varies in the attenuation region, we must study the behavior of the square root of the product and the quotient of these impedances. The allocation of poles and zeros for the latter functions may be obtained graphically from Fig. 2 if we recognize that the product of a pole and a zero results in nothing, just like the quotient of a pole and a pole or a zero and a zero. The product of two zeros is a zero squared, and of two poles, a pole squared; while the quotient of a pole and a zero is a pole squared, etc.

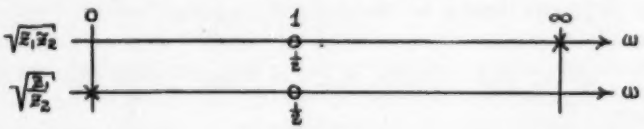


FIG. 3.

Fig. 3 illustrates this. The figures 1/2 under the zeros at ω_1 indicate that these are zeros to the one-half power. Note that $\sqrt{Z_1 Z_2}$ has no poles or zeros in the pass band, and $\sqrt{\frac{Z_1}{Z_2}}$ has no poles or zeros in the attenuation region. These are the regions respectively in which we wish to study these functions.

Analytically the above takes the following form:

$$\left.\begin{aligned} Z_1 &= jL_1\left(\frac{\omega^2 - \omega_1^2}{\omega}\right); \qquad \omega_1^2 = \frac{1}{L_1 C_1}, \\ Z_2 &= jL_2\omega, \end{aligned}\right\} \tag{6}$$

$$\sqrt{Z_1 Z_2} = \sqrt{L_1 L_2} \cdot \sqrt{\omega_1^2 - \omega^2}, \tag{7}$$

$$\sqrt{\frac{Z_1}{Z_2}} = \sqrt{\frac{L_1}{L_2}} \cdot \frac{\sqrt{\omega^2 - \omega_1^2}}{\omega}. \tag{8}$$

Function (7) is to be studied for $0 < \omega < \omega_1$, and function (8) for $\omega_1 < \omega < \infty$. The study of these two functions may be condensed into one problem by the following simple expedient. By factoring ω_1^2 out of the radical of (7) and ω^2 out of the radical of (8) we get:

$$\sqrt{Z_1 Z_2} = \sqrt{\frac{L_1}{C_2}} \cdot \sqrt{1 - \left(\frac{\omega}{\omega_1}\right)^2}; \qquad 0 < \left(\frac{\omega}{\omega_1}\right)^2 < 1, \tag{7a}$$

$$\sqrt{\frac{Z_1}{Z_2}} = \sqrt{\frac{L_1}{L_2}} \cdot \sqrt{1 - \left(\frac{\omega_1}{\omega}\right)^2}; \qquad 0 < \left(\frac{\omega_1}{\omega}\right)^2 < 1 \tag{8a}$$

If for (7*a*) we let:

$$x = \left(\frac{\omega}{\omega_1}\right)^2, \tag{9}$$

and for (8*a*) we let:

$$x = \left(\frac{\omega_1}{\omega}\right)^2, \tag{10}$$

then we see that (7*a*) and (8*a*) both reduce to the same form, namely:

$$F(x) = A\sqrt{1 - x}; \qquad 0 < x < 1, \tag{11}$$

which is quite simple. The constant A is, of course; different for the two functions. By the change of variable (10) we have converted the region *1 to* ∞, to *0 to 1*. The constant A we shall call the *nominal* value of the function. It is that value which (7*a*) assumes at the origin, and (8*a*) at infinity. Since (7*a*) is the characteristic impedance, its nominal value should in general be made equal to the value of resistance that the filter is to work out of and into. (8*a*) on the other hand is characteristic of the attenuation function. For the value unity, the attenuation is infinite. For a given departure from unity the attenuation is easily calculated by the relation (5). For example, if (8*a*)

departs from unity at some frequency by two percent, then the attenuation for that point is

$$\gamma_1 = ln\left(\frac{2}{.02}\right) = 4.6 \text{ napiers}$$

$$= 40 \text{ d.b.}$$

If the point of infinite attenuation is to fall at infinity, then the nominal value of (8*a*) should be taken as unity. If, further, the resistance we wish to work into and out of is R ohms, then we have

$$\left.\begin{aligned} \frac{L_1}{C_2} &= R^2, \\ L_1 &= L_2, \\ L_1C_1 &= \frac{1}{\omega_1^2}, \end{aligned}\right\} \tag{12}$$

and the design is complete. Note, however, that the point of infinite attenuation may be chosen to fall at any other point without changing the characteristic impedance function.

We have said nothing as yet about the general character of the function (11). We see by inspection that it is equal to its nominal value at zero, and is zero at the point $x = 1$. Between these points it has no maxima or minima but decreases in a monotonic manner. It should be remembered that the function (11) should be thought of as plotted against $\sqrt{x}$ and not x, because of the relations (9) and (10). In this way we obtain its normal behavior versus frequency. This behavior we find coincides precisely with that for the characteristic impedance function of the low-pass "constant-k" type filter. In fact, the design given by (12) will show the same behavior throughout as that for the corresponding "constant-k" design. The present illustration, based upon the impedances given by Fig. 2, is too simple a case to illustrate fully the flexibility of the Cauer method of design. Before leaving it, however, we wish to point out that identically the same result could have been obtained by using the reciprocal impedances to the above, i.e. by replacing

poles by zeros and vice versa. In that case the cut-off would be marked by a pole instead of a zero.

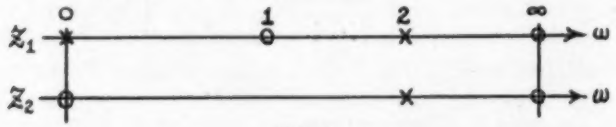


FIG. 4.

Consider the impedances specified by Fig. 4. Here we have added a pole to the right of the cut-off point. The characteristic impedance and attenuation functions are determined by the allocation of poles and zeros illustrated in Fig. 5. Here we see

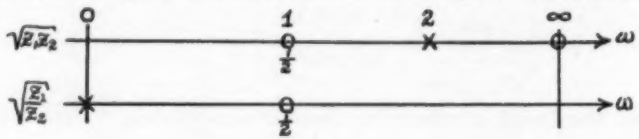


FIG. 5.

that the attenuation function has remained unchanged as compared to the previous case, but the characteristic impedance function has not. Analytically we have in this case:

$$Z_1 = P \cdot \frac{(\omega^2 - \omega_1^2)}{j\omega(\omega^2 - \omega_2^2)}, \tag{13}$$

$$Z_2 = Q \cdot \frac{\omega}{j(\omega^2 - \omega_2^2)}, \tag{14}$$

where P and Q are any constants. These functions are easily written down knowing the poles and zeros, and bearing in mind that the poles and zeros are simple. For further information on this point the reader is referred to the reference (4) given above.

The square roots of the product and quotient of these functions are

$$\sqrt{Z_1 Z_2} = \sqrt{PQ} \cdot \frac{\sqrt{\omega_1^2 - \omega^2}}{\omega_2^2 - \omega^2}, \tag{15}$$

$$\sqrt{\frac{Z_1}{Z_2}} = \sqrt{\frac{P}{Q}} \cdot \frac{\sqrt{\omega^2 - \omega_1^2}}{\omega}. \tag{16}$$

With the change of variable indicated by (9) and (10) respectively we have

$$\left.\begin{aligned}\sqrt{Z_1 Z_2} &= \frac{\sqrt{PQ}}{\omega_1} \cdot \frac{\sqrt{1-x}}{a_2 - x} \\ \sqrt{\frac{Z_1}{Z_2}} &= \sqrt{\frac{P}{Q}} \cdot \sqrt{1-x}\end{aligned}\right\} (0 < x < 1), \tag{17}$$

where the abbreviation was used:

$$a_2 = \left(\frac{\omega_2}{\omega_1}\right)^2. \tag{18}$$

The second of the functions (17) is identical in form with (11). The first is a new function. It contains an additional parameter a_2 which determines the location of the pole 2 in Fig. 4. The structure with impedances characterized by this figure will again be a low-pass filter. It will have the same attenuation properties for the same nominal value of the function (16) as the structure discussed above. The characteristic impedance function, however, will now be different. For a proper choice of a_2 it can be made to approximate its nominal value very much closer over a larger portion of the pass-band, than the function (11). It will be seen later that for $a_2 = 1.6$ this function approximates its nominal value to within 4 percent over about 86 percent of the pass band. In fact the function (15) is actually identical in form with that for the characteristic impedance of a properly composited low-pass Campbell type filter with single M-derived terminal transducers. The outstanding difference with the Cauer method lies in the fact that the attenuation function is not constrained to a definite behavior by this particular choice of characteristic impedance function. In our present example it is still the same as for the first example discussed above. It may be improved upon to any desired extent by choosing component impedances with more poles and zeros to the *left* of the frequency ω_1 in Fig. 4.

With these few examples the reader will probably be somewhat

oriented in the general way in which the Cauer method operates. At least he will have obtained a sufficient understanding of the method to appreciate its flexibilities. We wish now to discuss the method in a more general way, and to develop the necessary formulae and curves applicable to the design of four of the most common types of filters namely: Low-pass, high-pass, band-pass, and band-elimination, which we shall indicate by the abbreviations L.P., H.P., B.P., and B.E. respectively.

3. *Development of Design Formulae.* The reader will recognize from the foregoing that in the choice of the functions $\sqrt{Z_1Z_2}$ and $\sqrt{\frac{Z_1}{Z_2}}$ we are limited by the following specifications:

1. $\sqrt{Z_1Z_2}$ should have no poles or zeros in the pass-band.
2. $\sqrt{\frac{Z_1}{Z_2}}$ should have no poles or zeros in the attenuation bands.
3. The cut-off frequencies are marked by poles or zeros to the one-half power.
4. The balance of the poles and zeros must be simple and must separate each other.
5. $\sqrt{Z_1Z_2}$ must have a positive slope everywhere in the attenuation bands.
6. $\sqrt{\frac{Z_1}{Z_2}}$ must have a positive slope everywhere in the pass band.

When the functions are chosen according to these specifications, their product and quotient will always satisfy the necessary and sufficient conditions for physical realizability on the part of the component impedances. Any function of such a series may be chosen for the characteristic impedance independent of the particular choice which is made for the attenuation function and vice versa. The only restriction is that the functions shall, of course, belong to the same type, i.e. L.P., H.P., etc., whatever the case may be.

The series of functions which we shall formulate in this manner will contain certain parameters, like the a_2 of the first function

(17). The more complicated functions will contain several such parameters. These parameters definitely determine the tolerance and limits within which the given functions approximate their nominal values over their useful ranges. The outstanding problem to be solved is therefore the determination of the allowable values of these parameters which will lead to constant maximum tolerances, and a method whereby proper allowable values may be chosen to meet prescribed limits and tolerances. This is purely an algebraic problem. Its general solution may be obtained by the application of several mathematical methods one of which is known as the Tschebyscheff approximation method. This is the method which Cauer uses in his presentation. From the standpoint of generality and rigor it is perhaps the proper method to apply. However, to the average engineer such methods tend to obscure rather than clarify the general principles involved. The writer has found that, for those tolerances usually met in practice, the necessary functions are relatively simple, and for them rather elementary methods of analysis may be effectively applied. These methods involve the ordinary process of finding the maxima and minima of algebraic functions by equating the derivatives to zero. In the following we shall first formulate the series of functions which need be considered in connection with this method, and then study their properties in turn. Later we shall illustrate the use of the results so obtained in the design of filters for several typical cases.

In building this series of functions, it is not necessary to consider the attenuation and characteristic impedance functions separately. For example, for a low-pass filter, the characteristic impedance function is one which has poles and zeros located from the cut-off point to infinity. This, however, is also the case for the attenuation function for a high-pass filter. A similar situation holds with regard to the B.P. and B.E. types. Hence we can save time by considering these functions together. In order to help clarify matters, we accompany each analytical

function in turn by a figure which illustrates the corresponding distribution of poles and zeros for that case. We shall also see that by means of properly chosen approximations and changes of variable, we shall be able to transform all the functions which need be considered, into but six fundamental forms, so that in the end, only these need be studied. These six functions will be numbered consecutively by means of subscripts in the order of their ability to approximate their nominal values.

Another simplification may be made at this point. Instead of considering the actual component impedances Z_1 and Z_2, we define the *normalized* impedances:

$$z_1 = \frac{Z_1}{R}; \qquad z_2 = \frac{Z_2}{R}, \tag{19}$$

where R is the nominal value of the characteristic impedance, i.e. that value which the structure is intended to work into and out of. Thus:

$$\sqrt{Z_1 Z_2} = R\sqrt{z_1 z_2}. \tag{20}$$

In terms of these impedances z_1 and z_2, the nominal value of all functions becomes unity. Those component impedances for any desired nominal value are easily found from z_1 and z_2 by (19). This use of normalized impedances simplifies the functions somewhat.

Let us consider first the series of functions from which we may select for

$$\text{L.P. } \sqrt{z_1 z_2} \quad \text{or} \quad \text{H.P. } \sqrt{\frac{z_1}{z_2}}.$$

The first is represented graphically by Fig. 6. For this we have

FIG. 6.

$$\left.\begin{aligned} F(\omega) &= \frac{\sqrt{\omega_1^2 - \omega^2}}{\omega_1}; \\ 0 < \omega &< \omega_1 < \infty. \end{aligned}\right\} \tag{21}$$

Here let

$$\left(\frac{\omega}{\omega_1}\right)^2 = x. \tag{22}$$

Then (21) becomes

$$F_2(x) = \sqrt{1 - x}; \qquad (0 < x < 1). \tag{23}$$

Fig. 7 illustrates the next function in order:

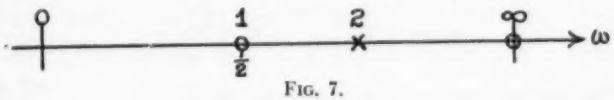


FIG. 7.

Here

$$\left.\begin{aligned} F(\omega) &= \frac{\omega_2^2\sqrt{\omega_1^2 - \omega^2}}{\omega_1(\omega_2^2 - \omega^2)}, \\ 0 < \omega < \omega_1 < \omega_2 < \infty. \end{aligned}\right\} \tag{24}$$

If we let

$$\left(\frac{\omega}{\omega_1}\right)^2 = x, \qquad \left(\frac{\omega_2}{\omega_1}\right)^2 = a_2, \tag{25}$$

equation (24) becomes:

$$F_4(x) = \frac{a_2\sqrt{1 - x}}{(a_2 - x)}, \qquad (0 < x < 1). \tag{26}$$

The next function is illustrated by Fig. 8. For this we have

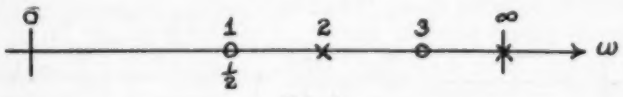


FIG. 8.

$$\left.\begin{aligned} F(\omega) &= \frac{\omega_2^2\sqrt{\omega_1^2 - \omega^2}(\omega_3^2 - \omega^2)}{\omega_1\omega_3^2(\omega_2^2 - \omega^2)}, \\ 0 < \omega < \omega_1 < \omega_2 < \omega_3 < \infty. \end{aligned}\right\} \tag{27}$$

Here we let

$$\left(\frac{\omega}{\omega_1}\right)^2 = x, \qquad \left(\frac{\omega_2}{\omega_1}\right)^2 = a_2, \qquad \left(\frac{\omega_3}{\omega_1}\right)^2 = a_3. \tag{28}$$

Then we have

$$F_6(x) = \frac{a_2\sqrt{1-x}(a_3-x)}{a_3(a_2-x)}, \qquad (0 < x < 1). \tag{29}$$

More complicated functions than F_6 are very rarely required in practice. Hence we shall not carry this series any farther, but rather turn to the corresponding series for

$$\text{H.P. } \sqrt{z_1 z_2} \quad \text{or} \quad \text{L.P. } \sqrt{\frac{z_1}{z_2}}.$$

The first of these is given in Fig. 9. Here

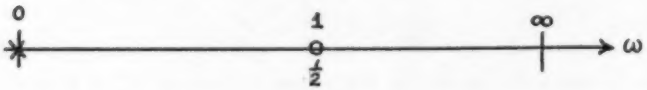


FIG. 9.

$$\left.\begin{aligned} F(\omega) &= \frac{\sqrt{\omega_1^2-\omega^2}}{j\omega}, \\ 0 < \omega_1 &< \omega < \infty. \end{aligned}\right\} \tag{30}$$

If we let

$$\left(\frac{\omega_1}{\omega}\right)^2 = x, \tag{31}$$

then (30) becomes F_2, eq. (23).

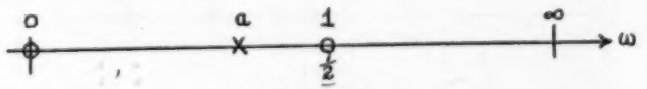


FIG. 10.

The next in this series is shown in Fig. 10. This is represented by the analytic expression:

$$\left.\begin{aligned} F(\omega) &= \frac{j\omega\sqrt{\omega_1^2-\omega^2}}{(\omega_a^2-\omega^2)}, \\ 0 < \omega_a &< \omega_1 < \omega < \infty. \end{aligned}\right\} \tag{32}$$

If we put

$$\left(\frac{\omega_1}{\omega}\right)^2 = x, \qquad \left(\frac{\omega_1}{\omega_a}\right)^2 = a_2, \tag{33}$$

then this function becomes F_4, eq. (26).

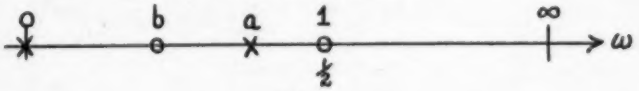


FIG. 11.

Finally the last in this series is illustrated in Fig. 11. For this we have:

$$\left.\begin{aligned} F(\omega) &= \frac{\sqrt{\omega_1^2 - \omega^2}(\omega_b^2 - \omega^2)}{j\omega(\omega_a^2 - \omega^2)}, \\ 0 < \omega_b < \omega_a < \omega_1 < \omega < \infty. & \end{aligned}\right\} \tag{34}$$

Here we put

$$\left(\frac{\omega_1}{\omega}\right)^2 = x, \qquad \left(\frac{\omega_1}{\omega_a}\right)^2 = a_2, \qquad \left(\frac{\omega_1}{\omega_b}\right)^2 = a_3, \tag{35}$$

and (34) becomes F_6, eq. (29).

Thus we see that the three functions F_2, F_4, and F_6 suffice to treat the L.P. and H.P. cases. In the treatment of the B.P. and B.E. types, we shall find these three functions again, plus three new ones. Before taking up the functions for these types we wish to make the following remarks relative to certain changes of variable which will be made use of.

In the case of B.P. and B.E. types, the distribution of poles and zeros for the characteristic impedance and attenuation functions as illustrated in Fig. 12 are considered to be sym-

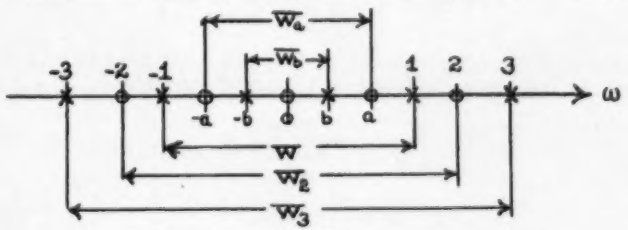


FIG. 12.

metrical about the frequency ω_0 which is located midway between the cut-off frequencies ω_{-1} and ω_1. We shall also introduce the notation

$$\left.\begin{aligned} w &= \omega_1 - \omega_{-1}, \\ w_2 &= \omega_2 - \omega_{-2}, \\ &\cdots\cdots \\ w_a &= \omega_a - \omega_{-a}, \\ w_b &= \omega_b - \omega_{-b}, \\ &\cdots\cdots \end{aligned}\right\} \tag{36}$$

to indicate the distances between corresponding poles and zeros to the right and left of ω_0. Here w indicates the band width.

At this point we introduce an approximation which holds very nearly for most practical cases, namely we assume that the width of the band is small compared to its distance from the origin. Analytically this is expressed by

$$\frac{w}{\omega_0} << 1 \tag{37}$$

with

$$\omega_0 = \frac{\omega_1 + \omega_{-1}}{2}. \tag{38}$$

By symmetry we also have

$$\left.\begin{aligned} \omega_1 &= \omega_0 + \frac{w}{2}; \qquad \omega_{-1} = \omega_0 - \frac{w}{2}; \\ \omega_2 &= \omega_0 + \frac{w_2}{2}; \qquad \omega_{-2} = \omega - \frac{w_2}{2}; \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \omega_a &= \omega_0 + \frac{w_a}{2}; \qquad \omega_{-a} = \omega_0 - \frac{w_a}{2}. \\ &\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned}\right\} \tag{39}$$

Now we introduce the following change of variable:

$$\omega = \omega_0 + \frac{w}{2}\Omega \tag{40}$$

or

$$\Omega = \frac{2(\omega - \omega_0)}{w}. \tag{41}$$

By (37), (39) and (40) we then have

$$\left.\begin{array}{ll} \omega^2 \leqq \omega_0^2 + w\omega_0\Omega, & \\ \omega_1^2 \leqq \omega_0^2 + w\omega_0; & \omega_{-1}^2 \leqq \omega_0^2 - w\omega_0, \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \\ \omega_a^2 \leqq \omega_0^2 + w_a\omega_0; & \omega_{-a}^2 \leqq \omega_0^2 - w_a\omega_0. \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \end{array}\right\} \tag{42}$$

So that

$$\left.\begin{array}{ll} \omega^2 - \omega_{-1}^2 \leqq w\omega_0(1+\Omega); & \omega_1^2 - \omega^2 \leqq w\omega_0(1-\Omega), \\ \omega^2 - \omega_{-2}^2 \leqq w\omega_0\left(\dfrac{w_2}{w} + \Omega\right); & \omega_2^2 - \omega^2 \leqq w\omega_0\left(\dfrac{w_2}{w} - \Omega\right), \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \\ \omega^2 - \omega_{-a}^2 \leqq w\omega_0\left(\dfrac{w_a}{w} + \Omega\right); & \omega_a^2 - \omega^2 \leqq w\omega_0\left(\dfrac{w_a}{w} - \Omega\right). \\ \cdots\cdots\cdots\cdots & \cdots\cdots\cdots\cdots \end{array}\right\} \tag{43}$$

The approximate substitutions (43) will be made use of in reducing the functions for the B.P. and B.E. types to simpler forms.

We consider first those functions for

$$\text{B.P. } \sqrt{z_1 z_2} \quad \text{or} \quad \text{B.E. } \sqrt{\frac{z_1}{z_2}}.$$

Fig. 13 illustrates the distribution of poles and zeros for the simplest of this series of functions. The cut-off frequencies

FIG. 13.

are here marked by a pole and zero respectively. The function of frequency which represents this case is given by

$$\left.\begin{array}{l} F(\omega) = \dfrac{j\omega\sqrt{\omega_1^2 - \omega^2}}{\omega_0\sqrt{\omega_{-1}^2 - \omega^2}} \leqq \sqrt{\dfrac{1-\Omega}{1+\Omega}}, \\ 0 < \omega_{-1} < \omega < \omega_1 < \infty, \end{array}\right\} \tag{44}$$

in which the substitutions (43) were used. Now replace

$$\Omega = x \tag{45}$$

and this becomes

$$F_1(x) = \sqrt{\frac{1-x}{1+x}}, \qquad (-1 < x < 1). \tag{46}$$

FIG. 14.

The next case is illustrated in Fig. 14. Here we have

$$\left.\begin{aligned} F(\omega) &= \frac{\sqrt{(\omega_{-1}^2 - \omega^2)(\omega_1^2 - \omega^2)}}{j\omega w} \leqq \sqrt{1-\Omega^2}, \\ 0 &< \omega_{-1} < \omega < \omega_1 < \infty. \end{aligned}\right\} \tag{47}$$

If we now put

$$\Omega^2 = x, \tag{48}$$

then this becomes F_2, eq. (23).

FIG. 15.

Next consider the distribution of poles and zeros given in Fig. 15. The function of frequency becomes

$$\left.\begin{aligned} F(\omega) &= \frac{\omega_0(\omega_{-2}^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{j\omega\sqrt{\omega_{-1}^2 - \omega^2}(\omega_2^2 - \omega^2)} \leqq \sqrt{\frac{1-\Omega}{1+\Omega}}\frac{\left(\frac{w_2}{w} + \Omega\right)}{\left(\frac{w_2}{w} - \Omega\right)}, \\ 0 &< \omega_{-2} < \omega_{-1} < \omega < \omega_1 < \omega_2 < \infty. \end{aligned}\right\} \tag{49}$$

Here let

$$\Omega = x, \qquad \frac{w_2}{w} = a_2. \tag{50}$$

Then (49) becomes

$$F_3(x) = \sqrt{\frac{1-x}{1+x}\frac{(a_2+x)}{(a_2-x)}}, \qquad (-1 < x < 1), \tag{51}$$

which is a function whose approximation ability places it between F_2 and F_4, as we shall see later.

FIG. 16.

The next case in order is illustrated by Fig. 16. For this figure we have

$$\left.\begin{aligned} F(\omega) &= \frac{j\omega w_2^2\sqrt{(\omega_{-1}^2-\omega^2)(\omega_1^2-\omega^2)}}{w(\omega_{-2}^2-\omega^2)(\omega_2^2-\omega^2)} \leqq \frac{\left(\frac{w_2}{w}\right)^2\sqrt{1-\Omega^2}}{\left(\frac{w_2}{w}\right)^2-\Omega^2}, \\ &0 < \omega_{-2} < \omega_{-1} < \omega < \omega_1 < \omega_2 < \infty. \end{aligned}\right\} \tag{52}$$

If we now let

$$\Omega^2 = x; \qquad \left(\frac{w_2}{w}\right)^2 = a_2, \tag{53}$$

then this becomes F_4, eq. (26).

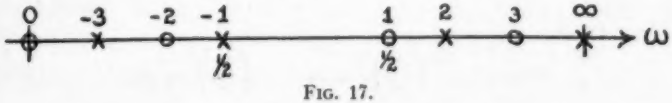


FIG. 17.

Next consider Fig. 17. For this distribution of poles and zeros we get

$$\left.\begin{aligned} F(\omega) &= \frac{j\omega(\omega_{-2}^2-\omega^2)\sqrt{\omega_1^2-\omega^2}(\omega_3^2-\omega^2)}{\omega_0(\omega_{-3}^2-\omega^2)\sqrt{\omega_{-1}^2-\omega^2}(\omega_2^2-\omega^2)} \\ &\leqq \sqrt{\frac{1-\Omega}{1+\Omega}\frac{\left(\frac{w_2}{w}+\Omega\right)\left(\frac{w_3}{w}-\Omega\right)}{\left(\frac{w_2}{w}-\Omega\right)\left(\frac{w_3}{w}+\Omega\right)}}, \\ &0 < \omega_{-3} < \omega_{-2} < \omega_{-1} < \omega < \omega_1 < \omega_2 < \omega_3 < \infty. \end{aligned}\right\} \tag{54}$$

Here we let

$$\Omega = x, \qquad \frac{w_2}{w} = a_2, \qquad \frac{w_3}{w} = a_3, \tag{55}$$

and (54) becomes

$$F_5(x) = \sqrt{\frac{1-x}{1+x}}\,\frac{(a_2+x)}{(a_2-x)}\,\frac{(a_3-x)}{(a_3+x)}, \qquad (-1 < x < 1). \tag{56}$$

FIG. 18.

The last function in this series is illustrated in Fig. 18. The analytic expression for it is given by

$$\left.\begin{aligned} F(\omega) &= \frac{w_2^2(\omega_{-3}^2-\omega^2)\sqrt{(\omega_{-1}^2-\omega^2)(\omega_1^2-\omega^2)}(\omega_3^2-\omega^2)}{j\omega w_3^2 w(\omega_{-2}^2-\omega^2)(\omega_2^2-\omega^2)} \\ &\cong \frac{w_2^2\sqrt{1-\Omega^2}\left(\left(\frac{w_3}{w}\right)^2-\Omega^2\right)}{w_3^2\left(\left(\frac{w_2}{w}\right)^2-\Omega^2\right)}, \\ 0 < \omega_{-3} < \omega_{-2} < \omega_{-1} < \omega < \omega_1 < \omega_2 < \omega_3 < \infty. \end{aligned}\right\} \tag{57}$$

Here we let

$$\Omega^2 = x, \qquad \left(\frac{w_2}{w}\right)^2 = a_2, \qquad \left(\frac{w_3}{w}\right)^2 = a_3, \tag{58}$$

and (57) becomes the function F_6, eq. (29).

We now proceed to develop the series of functions for:

$$\text{B.E. } \sqrt{z_1 z_2} \quad \text{or} \quad \text{B.P. } \sqrt{\frac{z_1}{z_2}},$$

FIG. 19.

The simplest of these is given in Fig. 19. Its analytic expression is:

$$\left.\begin{aligned} F(\omega) &= \sqrt{\frac{\omega_1^2 - \omega^2}{\omega_{-1}^2 - \omega^2}} \leqq \sqrt{\frac{\Omega - 1}{\Omega + 1}}, \\ 0 < \omega < \omega_{-1} < \omega_1 < \omega < \infty. \end{aligned}\right\} \tag{59}$$

If we let

$$\Omega = 1/x, \tag{60}$$

this becomes F_1, eq. (46).

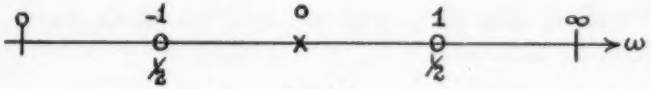


FIG. 20.

Next we have the function shown in Fig. 20. For this

$$\left.\begin{aligned} F(\omega) &= \frac{\sqrt{(\omega_{-1}^2 - \omega^2)(\omega_1^2 - \omega^2)}}{\omega_0^2 - \omega^2} \leqq \frac{\sqrt{\Omega^2 - 1}}{\Omega}, \\ 0 < \omega < \omega_{-1} < \omega_0 < \omega_1 < \omega < \infty. \end{aligned}\right\} \tag{61}$$

Here we put

$$\Omega^2 = 1/x, \tag{62}$$

and obtain F_2, eq. (23).

FIG. 21.

We then consider the function represented by the distribution of poles and zeros in Fig. 21, for which

$$\left.\begin{aligned} F(\omega) &= \frac{(\omega_{-a}^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{\sqrt{\omega_{-1}^2 - \omega^2}(\omega_a^2 - \omega^2)} \leqq \sqrt{\frac{\Omega - 1}{\Omega + 1}} \frac{\left(\Omega + \frac{w_a}{w}\right)}{\left(\Omega - \frac{w_a}{w}\right)}, \\ 0 < \omega < \omega_{-1} < \omega_{-a} < \omega_a < \omega_1 < \omega < \infty. \end{aligned}\right\} \tag{63}$$

Here we make the substitution:

$$\Omega = \frac{1}{x}; \qquad \frac{w}{w_a} = a_2 \tag{64}$$

and (63) becomes F_3, eq. (51).

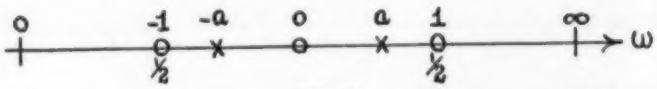


FIG. 22.

Now we form the analytic expression for the allocation of poles and zeros shown in Fig. 22, and obtain

$$\left.\begin{aligned} F(\omega) &= \frac{\sqrt{\omega_{-1}^2 - \omega^2}(\omega_0^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{(\omega_{-a}^2 - \omega^2)(\omega_a^2 - \omega^2)} \leqq \frac{\Omega\sqrt{\Omega^2 - 1}}{\Omega^2 - \left(\frac{w_a}{w}\right)^2}, \\ 0 &< \omega < \omega_{-1} < \omega_{-a} < \omega_0 < \omega_a < \omega_1 < \omega < \infty. \end{aligned}\right\} \tag{65}$$

For this function we let

$$\Omega^2 = 1/x, \qquad \left(\frac{w}{w_a}\right)^2 = a_2, \tag{66}$$

and find that it reduces to F_4, eq. (26).

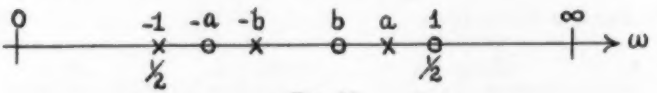


FIG. 23.

We consider next the function given in Fig. 23. Its analytic form is:

$$\left.\begin{aligned} F(\omega) &= \frac{(\omega_{-a}^2 - \omega^2)(\omega_b^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{\sqrt{\omega_{-1}^2 - \omega^2}(\omega_{-b}^2 - \omega^2)(\omega_a^2 - \omega^2)} \\ &\leqq \sqrt{\frac{\Omega - 1}{\Omega + 1}}\frac{\left(\Omega + \frac{w_a}{w}\right)\left(\Omega - \frac{w_b}{w}\right)}{\left(\Omega - \frac{w_a}{w}\right)\left(\Omega + \frac{w_b}{w}\right)}, \\ 0 &< \omega < \omega_{-1} < \omega_{-a} < \omega_{-b} < \omega_b < \omega_a < \omega_1 < \omega < \infty. \end{aligned}\right\} \tag{67}$$

Now we put

$$\Omega = 1/x, \qquad \frac{w}{w_a} = a_2, \qquad \frac{w}{w_b} = a_3, \tag{68}$$

and obtain F_8, eq. (56).

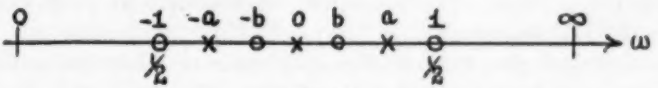


FIG. 24.

Finally, the last in this series is shown in Fig. 24. Its analytic representation is given by:

$$\left.\begin{aligned} F(\omega) &= \frac{\sqrt{\omega_{-1}^2 - \omega^2}(\omega_{-b}^2 - \omega^2)(\omega_b^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{(\omega_{-a}^2 - \omega^2)(\omega_0^2 - \omega^2)(\omega_a^2 - \omega^2)} \\ &\leqq \frac{\sqrt{\Omega^2 - 1}\left(\Omega^2 - \left(\frac{w_b}{w}\right)^2\right)}{\Omega\left(\Omega^2 - \left(\frac{w_a}{w}\right)^2\right)}, \\ 0 < \omega < \omega_{-1} < \omega_{-a} < \omega_{-b} < \omega_0 < \omega_b < \omega_a < \omega_1 < \omega < \infty. \end{aligned}\right\} \tag{69}$$

If we make the substitution here that

$$\Omega^2 = 1/x, \qquad \left(\frac{w}{w_a}\right)^2 = a_2, \qquad \left(\frac{w}{w_b}\right)^2 = a_3, \tag{70}$$

we find that (69) becomes F_6, eq. (29).

At this point the reader will perhaps feel that the present method of filter design runs quite heavily to formulae, so heavy in fact that the fundamental principles of the method are threatened with obscurity. However, this is an inherent characteristic of network theory anyway, especially in the field of network synthesis. Here, as in all other branches of this general subject, the student must at the outset carry on in a temporary fog with the hope that clarity will be eventually gained through perseverance. It is a well known characteristic of advanced network theory that relationships, which at first acquaintance seem very formidable, become relatively elementary after some

familiarity with them has been gained. So it is with the above manipulations. Once the fundamental idea of the formation of these functions has been understood, their formidable appearance disappears, especially since the end results are so relatively simple as those embodied in the six functions to which our entire system reduces.

Let us repeat these six functions again in order, before we take up a closer study of their individual properties. They are

$$F_1(x) = \sqrt{\frac{1-x}{1+x}}, \qquad (-1 < x < 1), \tag{71}$$

$$F_2(x) = \sqrt{1-x}, \qquad (0 < x < 1), \tag{72}$$

$$F_3(x) = \sqrt{\frac{1-x}{1+x}}\,\frac{(a_2+x)}{(a_2-x)}, \qquad (-1 < x < 1), \tag{73}$$

$$F_4(x) = \frac{a_2\sqrt{1-x}}{(a_2-x)}, \qquad (0 < x < 1), \tag{74}$$

$$F_5(x) = \sqrt{\frac{1-x}{1+x}}\,\frac{(a_2+x)(a_3-x)}{(a_2-x)(a_3+x)}, \qquad (-1 < x < 1), \tag{75}$$

$$F_6(x) = \frac{a_2\sqrt{1-x}(a_3-x)}{a_3(a_2-x)}, \qquad (0 < x < 1). \tag{76}$$

First, it should be remembered that any one of these six functions may be chosen for the attenuation function $\sqrt{\frac{z_1}{z_2}}$ independently of the choice which is made for the characteristic impedance function $\sqrt{z_1 z_2}$. When this choice has been made, and the parameters determined in accordance with prescribed limits and tolerances, the functions must be converted into the functions to which they are equivalent, according to the type of filter (L.P., H.P., etc.) to which they are supposed to belong. Then, by forming the product and quotient of these converted functions, we obtain the analytic expressions for z_1 and z_2, from which Z_1 and Z_2 may be found by multiplying by the value of resistance into which and out of which the section is to operate, as expressed by the relation (19). Finally, the impedances Z_1

and Z_2 are realized physically according to known methods. It should also be remembered that for the L.P. and H.P. types only F_2, F_4, or F_6 may be chosen, since these are the only ones which were obtained for these types. For the B.P. and B.E. types, any of the six functions may be chosen.

In designating filters designed by this method, the term "type" should be used to designate the kind of filter, i.e. whether L.P. or H.P., etc. The term "class" should refer to the particular kind of functions which are chosen for its realization. For example, a B.P. type, class 35 designates that F_3 has been chosen for the characteristic impedance and F_5 for the attenuation function. Note that the first figure refers to the impedance function and the second to the attenuation function. We now proceed with the discussion of the various functions themselves.

The function F_1, which may be chosen for B.P. and B.E. types

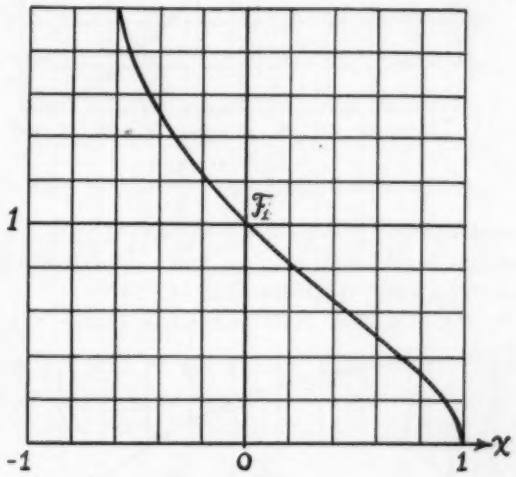


FIG. 25.

only, is illustrated by the Fig. 25. The origin corresponds to the center of the band, and the points $\pm$ 1 to the cut-off points. It is an odd function having a 45° slope at the point $x = 0$. It

approximates unity very poorly. Consequently it would hardly ever be chosen except for the fact that it may lead to a very simple structure. If we define the "percent tolerance" as the maximum percentage deviation from unity, and the "percent

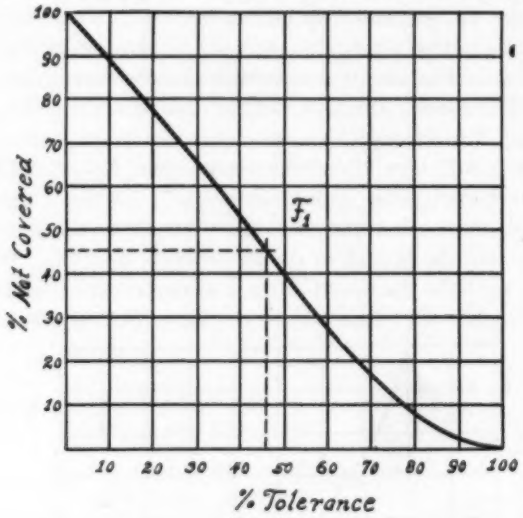


FIG. 26.

not covered" as that portion of the region -1 to $+1$ in which this tolerance is exceeded, then we find a relationship between these quantities which is illustrated in Fig. 26.

For F_2 the situation looks somewhat better. Figure 27

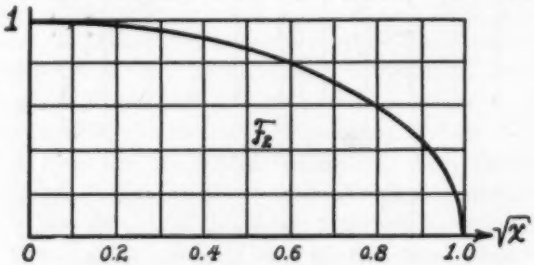


FIG. 27.

illustrates this. As mentioned before, this function is identical with that for the characteristic impedance of a "constant-k" Campbell type filter. The function is an even one. Hence the behavior for the region -1 to 0 is the image of that shown. For a L.P. type, the origin remains the origin. For a H.P. type, it represents infinity. For a B.P. or B.E. type it corresponds to the center of the band. The relation between percent tolerance and percent not covered is given in Fig. 28.

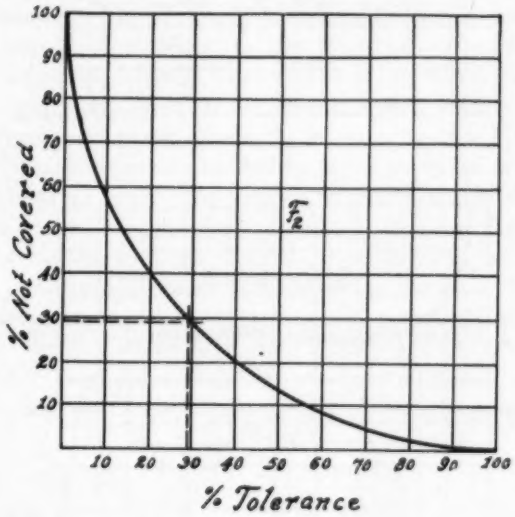


FIG. 28.

F_3 is the first function which contains an arbitrary parameter. Hence its behavior is not uniquely fixed, but may be altered by choosing various values for this parameter. In Figs. 29 and 30 are shown two such behaviors. We see that it is again an odd function. In order to be able to tell whether or not this function is capable of meeting prescribed requirements, it is necessary that we develop a relationship between the arbitrary parameter a_2 and the quantities: *Percent tolerance* and *percent not covered*.

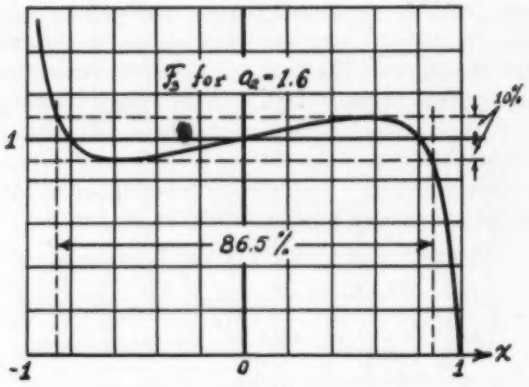


FIG. 29.

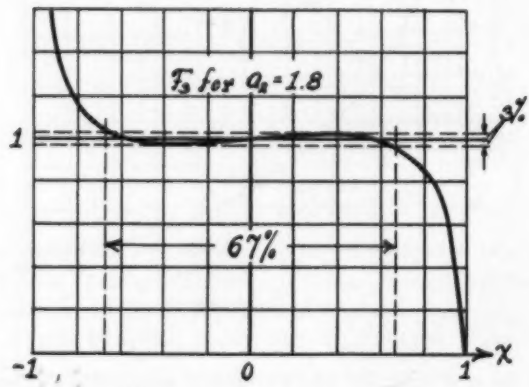


FIG. 30.

This is done by the simple algebraic method of determining the location and magnitude of the maxima and minima of this function. Equating the derivative of F_3 to zero we find:

$$x^2 = \frac{a_2(2 - a_2)}{2a_2 - 1}. \tag{77}$$

In order that this equation may have real roots between -1 and $+1$, it is necessary that

$$0 < \frac{a_2(2 - a_2)}{2a_2 - 1} < 1,$$

from which we get

$$1 < a_2 < 2. \tag{78}$$

Thus the choice of possible values for this parameter is considerably restricted. By choosing various values for a_2 and calculating corresponding deviations from unity as well as the limits within which this deviation is not exceeded, we obtain data from which the curves of Fig. 31 are drawn. These show again the fact illustrated by the curves for F_1 and F_2, that decreased tolerance is obtained only at the expense of a smaller percentage of the band covered. The figures 29 and 30 should be again examined in order to appreciate this point. However, the general result obtainable with this function begins to look pretty good. When used as a characteristic impedance function, it is an appreciable improvement upon F_2 which represents the "constant-k" design. When used as an attenuation function, it will give rise to three points of infinite attenuation, namely at those points where the function equals unity. These correspond to

$$\left.\begin{aligned} &x = 0 \\ \text{and} \qquad &x = \pm \sqrt{a_2(2 - a_2)}. \end{aligned}\right\} \tag{79}$$

For the case of Fig. 30 the attenuation will never go below the value:

$$\gamma_1 = 20 \log \left(\frac{1.97}{.03}\right) = 36.4 \text{ d.b.}$$

However, this value will be reached at some distance from cut-off, namely, at the frequencies corresponding to the points $x = \pm .67$. Hence this function will not give rise to exceptionally sharp cut-off properties, although they will be somewhat better than

those obtained with a "constant-k" design. For the case of Fig. 29, the cut-off will be considerably sharper, but the minimum attenuation lower, namely,

$$\gamma_1 = 20 \log \left(\frac{1.9}{.1}\right) = 25.6 \text{ d.b.}$$

By the use of Fig. 31, we can easily determine whether or not this function is able to meet the requirements set forth in a particular case.

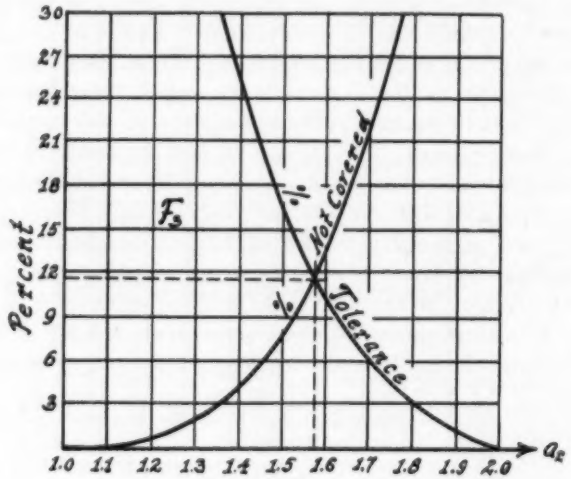


FIG. 31.

We now turn to function F_4, eq. (74). This also contains a parameter a_2. Hence various tolerances and limits may be obtained with it, as in the case of F_3. For F_4, the percent tolerance for a given percent not covered, will be less. Figures 32 and 33 illustrate two possible results which may be obtained by this function. In order to investigate all possibilities, we study the maximum and minimum properties of F_4 in the usual manner.

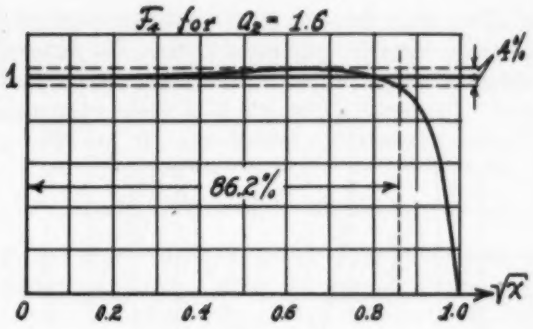


FIG. 32.

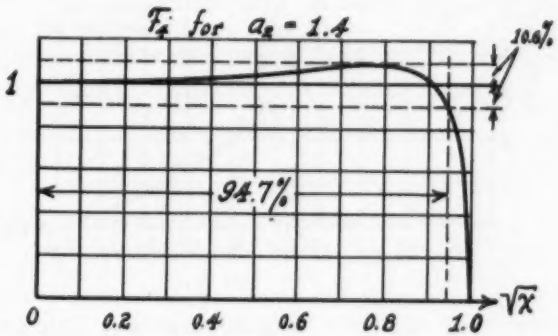


FIG. 33.

Setting the derivative equal to zero gives

$$x = 2 - a_2. \tag{80}$$

In order that this shall be a real value between zero and unity, we must have

$$1 < a_2 < 2. \tag{81}$$

Again, by assigning various values to a_2 and calculating the corresponding tolerances and portions of the band in which these are exceeded, we obtain data from which to plot the curves of

Fig. 34. They have the same general shape as those for F_3, but exhibit smaller tolerances and wider limits. The figures 32 and 33 should be re-examined for two possible cases. The first of these shows the usual choice which is made when using this function as a characteristic impedance. It has been already

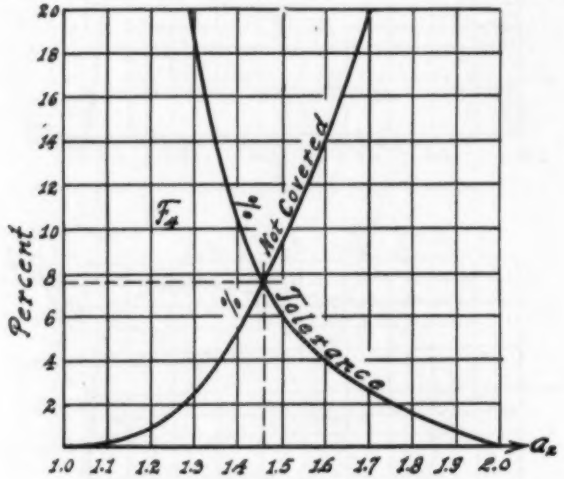


FIG. 34.

mentioned that this function is identical with that for the characteristic impedance of a properly composited Campbell type filter with single M-derived terminal transducers. In comparing the latter with our function F_4, we should note that the following relation holds between Zobel's parameter m and our a_2

$$a_2 = \frac{1}{1 - m^2}. \tag{82}$$

The figure 32 for $a_2 = 1.6$ therefore corresponds to $m = 0.6125$. The usual choice for m is 0.6. The figure 34 enables us to see at a glance what F_4 is capable of doing in the way of meeting the requirements of a specific case.

It may also be interesting as well as useful to determine those

values of x for which F_4 becomes equal to unity, which are the points at which infinite attenuation will be had when this function is used as an attenuation function. We find that these points are given by

$$\left.\begin{aligned} x &= 0 \\ \text{and} \qquad x &= a_2(2 - a_2). \end{aligned}\right\} \tag{83}$$

For L.P. and H.P. types the first of these corresponds to zero and infinity respectively, while the second locates one other point near the cut-off frequency. For B.P. and B.E. types the point $x = 0$ corresponds to the center of the band, while the second relation (83) determines two symmetrically located points near the cut-off frequencies.

Let us now study the function F_5 of eq. (75). Setting the derivative equal to zero gives

$$Ax^4 - 2Bx^2 + C = 0, \tag{84}$$

where

$$\left.\begin{aligned} A &= 2(a_3 - a_2) + 1, \\ B &= \tfrac{1}{2}(a_2^2 + a_3^2) - (a_3 - a_2)(a_2a_3 - 1), \\ C &= 2a_2a_3(a_2 - a_3 + \tfrac{1}{2}a_2a_3). \end{aligned}\right\} \tag{85}$$

Hence

$$x^2 = \frac{B}{A}\left[1 \pm \sqrt{1 - \frac{AC}{B^2}}\right]. \tag{86}$$

In order that x^2 may have positive roots between zero and unity we must have

$$0 \leqq 2B \leqq A; \qquad 0 \leqq AC \leqq B^2. \tag{87}$$

Since it is necessary that

$$a_3 \geqq a_2,$$

it follows that

$$A \geqq 0,$$

so that the inequality

$$AC \geqq 0$$

becomes simply

$$C \geqq 0,$$

which gives

$$a_2 - a_3 + \tfrac{1}{2}a_2a_3 \geqq 0$$

or

$$a_3 \leqq \frac{2a_2}{2 - a_2}. \tag{88}$$

This gives us an upper limit for a_3 in terms of a_2. Investigation shows that the inequality (88) insures

$$2B \geqq 0.$$

Hence, besides (88) we must also see to it that

$$2B \leqq A$$

and

$$AC \leqq B^2.$$

The first of these gives

$$(2a_2 - 1)a_3^2 - 2a_2^2a_3 - (a_2^2 - 1) \geqq 0. \tag{89}$$

The second gives

$$(2a_2 - 1)^2a_3^3 - (4a_2^3 + 8a_2^2 - 9a_2 - 4)a_3^2 - (4a_2^3 + 9a_2^2 - 8a_2 - 4)a_3 - a_2(2 - a_2)^2 \geqq 0. \tag{90}$$

Here we find that the conditions (90) and (88) always insure (89). (90) determines the upper limit for a_3 in terms of a_2. Thus the relations (88) and (90) together insure the fulfillment of the inequalities (87).

In figure 35 are shown the plots of the relations (88) and (90) versus a_2. The point at which the upper limit (88) meets the lower limit (90) is obtained by equating these relations. This gives

$$a_2^2 + 2a_2 - 4 = 0$$

or

$$a_2 = -1 + \sqrt{5} = 1.235,$$

so that we see that

$$1 < a_2 < 1.235. \tag{91}$$

The corresponding limit for a_3 is found by substituting the

upper limit (91) back into either (88) or (90). We find

$$a_3 = 3.23 \quad \text{for} \quad a_2 = 1.235. \tag{92}$$

This locates the upper right-hand point in Fig. 35. This figure shows that the choice of the parameters a_2 and a_3 which gives rise to maxima and minima for the function F_5 within the region $-1 < x < 1$ is very much restricted.

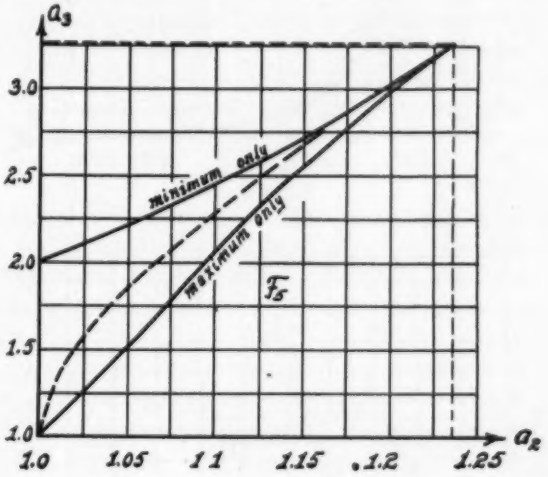


FIG. 35.

Our problem requires that only such values for these parameters be chosen for which the maxima and minima of the function correspond to equal deviations from the value unity. The dotted curve of Fig. 35 corresponds to such values. By choosing a number of points on this dotted curve and calculating the corresponding tolerances and portions of the band in which these are exceeded, we obtain data from which to plot the curves of Figs. 36 and 37. By means of these we are able to see at a glance what the possibilities of this function are, and what values for a_2 and a_3 should be chosen in order to meet a given

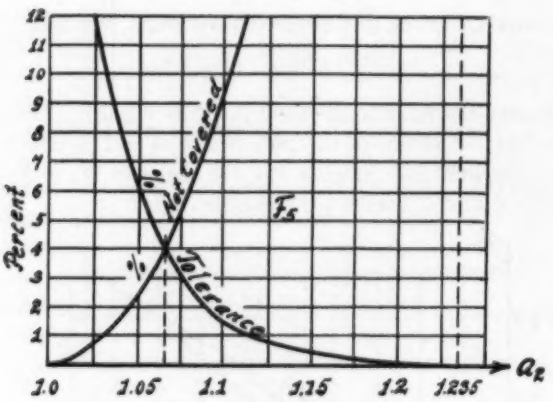


FIG. 36.

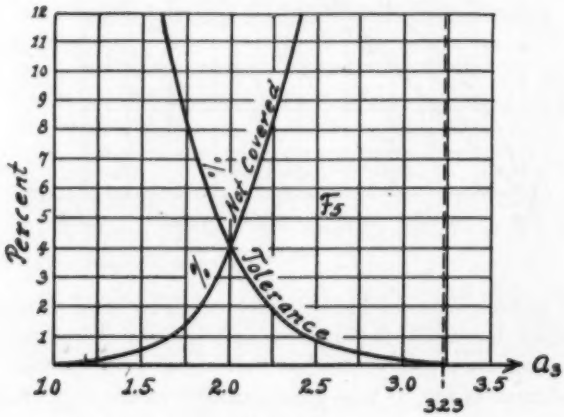


FIG. 37.

tolerance and coverage which may be obtained by this function for a specific case. Again we see that smaller tolerances and wider limits may be had than with the preceding function. Fig. 38 illustrates one possible result.

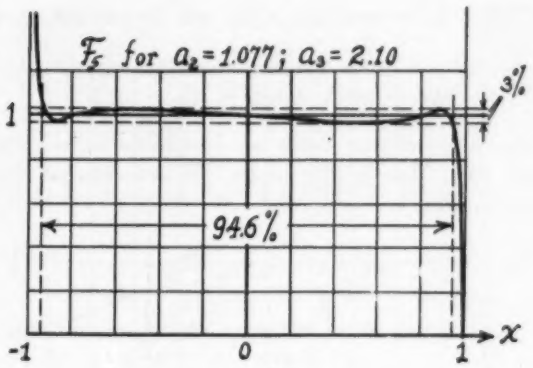


FIG. 38.

The function is again odd. It passes through unity at five points, which will correspond to points of infinite attenuation when this function is used as an attenuation function. These points are given by

$$\left.\begin{aligned} &x = 0 \\ &\text{and} \\ &x^2 = (a_2a_3 - a_3 + a_2) \\ &\quad - \frac{(a_3 - a_2)^2}{2}\left[1 \pm \sqrt{1 - \frac{4(a_2a_3 - a_3 + a_2 - 1)}{(a_3 - a_2)^2}}\right]. \end{aligned}\right\} \tag{93}$$

By means of these relations, the points of infinite attenuation may be predicted.

Finally, we study the function F_6 of eq. (76). This is again an even function. Setting the derivative equal to zero gives

$$x = \frac{3a_2 - a_3}{2}\left[1 \pm \sqrt{1 - \frac{4a_2a_3 - 8(a_3 - a_2)}{(3a_2 - a_3)^2}}\right]. \tag{94}$$

Real roots between 0 and 1, therefore, exist only when

$$\left.\begin{aligned} &0 < (3a_2 - a_3) < 1 \\ &\text{and} \\ &0 < \frac{4a_2a_3 - 8(a_3 - a_2)}{(3a_2 - a_3)^2} < 1. \end{aligned}\right\} \tag{95}$$

The second of these conditions splits into the two following:

$$\begin{aligned} a_2 a_3 &> 2(a_3 - a_2) \\ (3a_2 - a_3)^2 &> 4[a_2 a_3 - 2(a_3 - a_2)]. \end{aligned} \tag{96}$$

Examination shows that these two conditions include the first condition (95). Hence (96) states the case completely. We find that it results in

$$\left.\begin{aligned} 9a_2 - 8 < a_3 < \frac{2a_2}{2 - a_2} \\ \text{and} \\ 1 < a_2 < \frac{4}{3}. \end{aligned}\right\} \tag{97}$$

The second of these is found by locating that point at which the upper and lower limits for a_3 meet, i.e. from

$$\frac{2a_2}{2 - a_2} = 9a_2 - 8. \tag{98}$$

The corresponding upper limit for a_3 is found by substituting that for a_2 back into either of the expressions (98). We have

$$a_3 = 4 \quad \text{for} \quad a_2 = \frac{4}{3}. \tag{99}$$

The relations (97) and the upper limits (99) are illustrated in the figure 39, which shows that only those values for a_2 and a_3 which correspond to points within the area enclosed by the two curves will give rise to maxima and minima of the function within the limits: $0 < x < 1$. Again we wish to choose only those parameter values for which the maxima and minima correspond to equal deviations from the value unity. Such values are found for points located on the dotted curve of Fig. 39.

By choosing a number of such points and calculating the corresponding tolerances and portions of the band in which these are exceeded, we obtain data for the curves of Figs. 40 and 41,

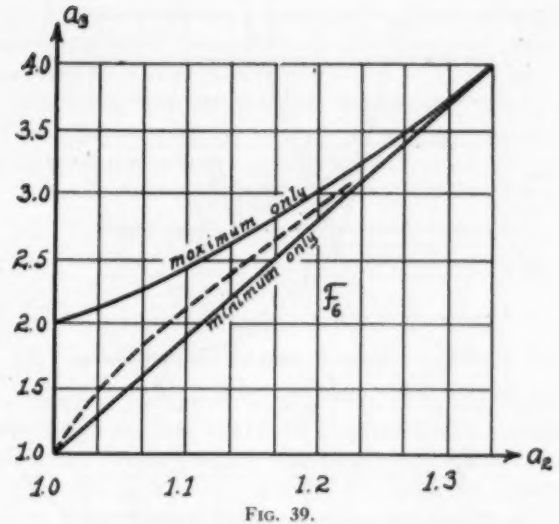


FIG. 39.

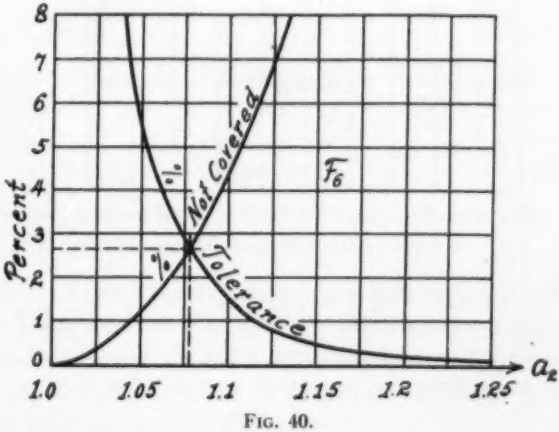


FIG. 40.

which show the possibilities of this function for approximation purposes. We see that the tolerances are in general smaller

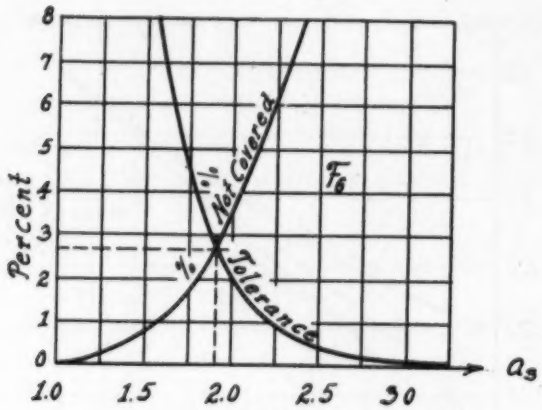


FIG. 41.

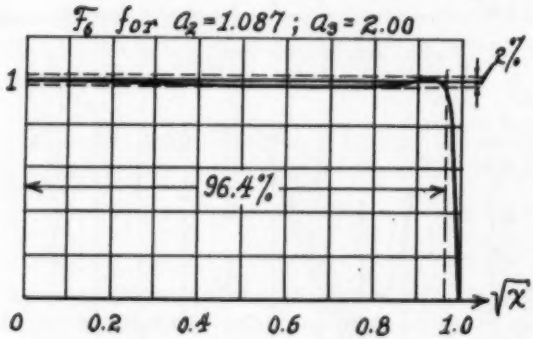


FIG. 42.

and the corresponding coverage larger than for F_5. Fig. 42 illustrates a possible result which may be obtained with F_6.

When used as an attenuation function, F_6 will give rise to infinite attenuation at those points corresponding to

$$\left.\begin{aligned} x &= 0, \\ \text{and} \\ x &= \frac{a_2^2(2a_3+1)-a_3^2}{2a_2^2}\left[1 \pm \sqrt{1-\frac{4a_2^3a_3(a_2a_3+2a_2-2a_3)}{[a_2^2(2a_3+1)-a_3^2]^2}}\right]. \end{aligned}\right\} \quad (100)$$

It is interesting to note that this function corresponds in general to that which Zobel obtains for a characteristic impedance by means of his double M-derived method.

In concluding this study of the functions F_1 to F_6, let us point out the following facts worth noting particularly. The possibilities of the various functions are best illustrated by

Fig. 26 for F_1,
Fig. 28 for F_2,
Fig. 31 for F_3,
Fig. 34 for F_4,
Figs. 36 and 37 for F_5,
Figs. 40 and 41 for F_6.

As an index of the overall or ultimate ability of a function in meeting tolerances and limits, we may specify that percentage for which the tolerance equals the part of the band not covered. Except for the functions F_1 and F_2, this point is given by the intersection of the two curves. In order, these percentages are

$$\left.\begin{array}{r} 45.5 \text{ for } F_1, \\ 29.0 \text{ for } F_2, \\ 11.56 \text{ for } F_3, \\ 7.7 \text{ for } F_4, \\ 4.0 \text{ for } F_5, \\ 2.66 \text{ for } F_6, \end{array}\right\} \tag{101}$$

It is rather interesting to note that these figures for the various functions bear the approximate ratios indicated by

$$\frac{F_1}{F_2} \cong \frac{F_3}{F_4} \cong \frac{F_5}{F_6} \cong \frac{3}{2}. \tag{102}$$

Whether or not this has any deeper significance we do not know. However, the relative values of the series of functions are characteristically indicated by (101).

It is also worth noting that, whereas the functions F_2, F_4, F_6 are identical in form with those which Zobel finds for the char-

acteristic impedance of the "constant-k," single and double M-derived types respectively, the functions F_1, F_3, F_5 are new, and exhibit properties which place their ultimate values between those of the even functions. Thus, whereas the Zobel method of design gives us no intermediate choice between the "constant-k" and the single M-derived types, for example, our present method does, namely in terms of F_3. The same applies to F_5 as an intermediate choice between functions equivalent to the single and double M-derived types. This, of course, applies only to B.P. and B.E. types, since, as already pointed out, the odd functions F_1, F_3, F_5 cannot be chosen for L.P. and H.P. types.

It should be remembered that in the present method, these functions may also be independently chosen for the attenuation function of a filter. Just how this material, which has been presented so far, may be used in a design will now be taken up.

4. *Applications to Specific Cases.* Suppose it is required to design a low-pass filter to meet the following specifications:

$$R = 500 \text{ ohms},$$
$$\omega_1 = 2\pi \cdot 3000.$$

In addition, the attenuation is to be at least 34 d.b. at $2\pi \cdot 3500$ cycles/sec. and should not fall below this value at any other point. The characteristic impedance is to approximate the nominal value to within 2 percent over 95 percent of the pass band.

Let us meet the attenuation requirements first. The change of variable which applies to L.P. $\sqrt{\frac{z_1}{z_2}}$, is given by (31), or

$$\frac{\omega_1}{\omega} = \sqrt{x}$$

Hence the point $\omega = 2\pi \cdot 3500$ corresponds to

$$\frac{3000}{3500} = \sqrt{x} = 0.857.$$

This means that the function is to cover 85.7 percent of the band. The allowable tolerance is determined from the minimum attenuation. Here we see that

$$20 \log \left(\frac{2}{.04}\right) = 34 \text{ d.b.}$$

Hence the tolerance is 4 percent, and the part of the band not covered is 14.3 percent. Fig. 34 shows that F_4 meets these requirements for

$$a_2 = 1.6.$$

Hence we see by eq. (32) that for this case

$$\sqrt{\frac{z_1}{z_2}} = \frac{j\omega\sqrt{\omega_1^2 - \omega^2}}{(\omega_a^2 - \omega^2)}, \tag{103}$$

where by (33)

$$\left(\frac{\omega_1}{\omega_a}\right)^2 = 1.6, \qquad \omega_a = 2\pi \cdot 2370. \tag{104}$$

The characteristic impedance requirements of 2 percent over 95 percent are found from figures 40 and 41 to be met by F_6 for

$$a_2 = 1.087, \quad a_3 = 2.00.$$

Equation (27) which is the equivalent of F_6 for this case, therefore determines

$$\sqrt{z_1 z_2} = \frac{\omega_2^2\sqrt{\omega_1^2 - \omega^2}(\omega_3^2 - \omega^2)}{\omega_1\omega_3^2(\omega_2^2 - \omega^2)}, \tag{105}$$

where from (28)

$$\left(\frac{\omega_2}{\omega_1}\right)^2 = 1.087, \qquad \left(\frac{\omega_3}{\omega_1}\right)^2 = 2.00,$$

$$\omega_2 = 2\pi \cdot 3130, \qquad \omega_3 = 2\pi \cdot 4240. \tag{106}$$

The product and quotient of (105) and (103) give z_1 and z_2 respectively. By (19), the component impedances are therefore

$$Z_1 = \frac{R\omega_2^2}{\omega_1\omega_3^2} \cdot \frac{j\omega(\omega_1^2 - \omega^2)(\omega_3^2 - \omega^2)}{(\omega_a^2 - \omega^2)(\omega_2^2 - \omega^2)} \tag{107}$$

and

$$Z_2 = \frac{R\omega_2^2}{\omega_1\omega_3^2} \cdot \frac{(\omega_a^2 - \omega^2)(\omega_3^2 - \omega^2)}{j\omega(\omega_2^2 - \omega^2)}, \tag{108}$$

Substituting numerical values we find

$$\frac{R\omega_2^2}{\omega_1\omega_3^2} = 0.0144. \tag{109}$$

The allocation of poles and zeros for the impedance functions (107) and (108) is given in Fig. 43. One form of physical structure for each impedance is given in Fig. 44. The values of inductance and capacitance in these structures may be determined from (104), (106), (107), and (108) by Foster's method referred to in footnote (4) above.

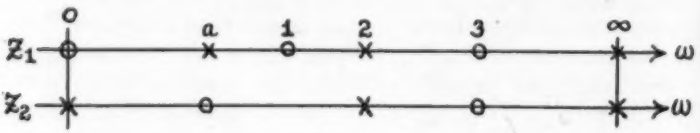


FIG. 43.

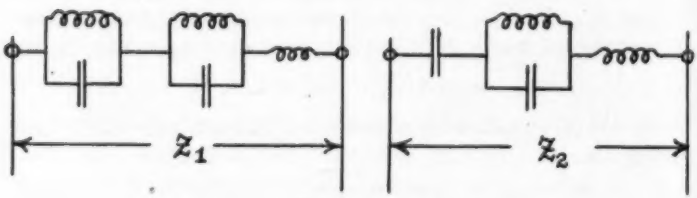


FIG. 44.

Let us illustrate with one more case. Suppose it is desired to design a band-pass filter to meet the following specifications:

$$\omega_0 = 2\pi \cdot 10^5, \qquad w = 2\pi \cdot 10^4, \tag{110}$$

i.e. a band width of 10,000 cycles/sec., with the center located at 100,000 cycles/sec. Further, we may wish to work into and out of say 10,000 ohms, i.e.

$$R = 10^4. \tag{111}$$

The characteristic impedance need only equal this nominal value

to within 8 percent over 80 percent of the pass band. The attenuation, however, shall be at least 46 d.b. (this corresponds to a numerical ratio of 200 to 1) at points 1000 cycles/sec. above and below cut-off and shall not drop below this value at any other point.

The attenuation of 46 d.b., we see, corresponds to a tolerance of one percent, since

$$20 \log \left(\frac{1 + .99}{1 - .99} \right) \leqq 46.$$

The point 1000 cycles/sec., above or below cut-off is found by eq. (41) to correspond to

$$\Omega = \frac{2 \cdot 2\pi \cdot 6000}{2\pi \cdot 10{,}000} = 1.2.$$

Under the development of $\sqrt{\frac{z_1}{z_2}}$ – functions for B.P. types, we employed the further change of variable

$$\frac{1}{\Omega} = x \text{ for odd functions,}$$

$$\frac{1}{\Omega} = \sqrt{x} \text{ for even functions}$$

Hence we see that, whether we choose an odd or even function, the requirements are that it shall give rise to a tolerance of one percent over:

$$\frac{100}{1.2} = 83.3 \text{ percent}$$

of the band, or with 16.7 percent of the band not covered. From Fig. 34 we see that F_4 does not meet these requirements. However, figures 36 and 37 show that F_5 will easily meet them with

$$a_2 = 1.125, \qquad a_3 = 2.5. \tag{112}$$

Fig. 35 shows that these values lie on the dotted line as required. This serves as a check on the reading of points in figures 36 and

37. By exterpolating one of the curves in the latter, we see that the values (112) for F_5 will give rise to a tolerance of 0.82 percent over 85 percent of the band, or a minimum attenuation of 47.6 d.b. which will be reached at points 880 cycles/sec. above and below the cut-off frequencies. (Note that 47.6 d.b. corresponds to a numerical ratio of 244 to 1.)

The tolerance and limits for the characteristic impedance function may easily be met by F_3, as seen from Fig. 31. If for this function we choose

$$a_2 = 1.67, \tag{113}$$

we see that we obtain a 7 percent tolerance over 81 percent of the band.

The characteristic impedance function is found from equations (49) and (50). The latter, together with (113) gives

$$w_2 = wa_2 = 2\pi \cdot 1.67 \cdot 10^4,$$

so that

$$\omega_2 = \omega_0 + \frac{w_2}{2}, \qquad \omega_{-2} = \omega_0 - \frac{w_2}{2},$$

become

$$\left.\begin{aligned} \omega_2 &= 2\pi \cdot 1.0835 \cdot 10^5, \\ \omega_{-2} &= 2\pi \cdot 0.9165 \cdot 10^5. \end{aligned}\right\} \tag{114}$$

Equation (49) gives

$$\sqrt{z_1 z_2} = \frac{\omega_0(\omega_{-2}^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{j\omega\sqrt{\omega_{-1}^2 - \omega^2}(\omega_2^2 - \omega^2)}. \tag{115}$$

For the attenuation function we turn to equations (67) and (68). The latter, together with (112), gives

$$\left.\begin{aligned} w_a &= \frac{w}{a_2} = 2\pi \cdot 0.889 \cdot 10^4, \\ w_b &= \frac{w}{a_3} = 2\pi \cdot 0.4 \cdot 10^4, \end{aligned}\right\} \tag{116}$$

so that

$$\left.\begin{aligned} \omega_a &= \omega_0 + \frac{w_a}{2} = 2\pi \cdot 1.0445 \cdot 10^5, \\ \omega_{-a} &= \omega_0 - \frac{w_a}{2} = 2\pi \cdot 0.9555 \cdot 10^5, \\ \omega_b &= \omega_0 + \frac{w_b}{2} = 2\pi \cdot 1.02 \cdot 10^5, \\ \omega_{-b} &= \omega_0 - \frac{w_b}{2} = 2\pi \cdot 0.98 \cdot 10^5. \end{aligned}\right\} \tag{117}$$

Equation (67) gives

$$\sqrt{\frac{z_1}{z_2}} = \frac{(\omega_{-a}^2 - \omega^2)(\omega_b^2 - \omega^2)\sqrt{\omega_1^2 - \omega^2}}{\sqrt{\omega_{-1}^2 - \omega^2}(\omega_{-b}^2 - \omega^2)(\omega_a^2 - \omega^2)}. \tag{118}$$

R times the product and quotient of (115) and (118) respectively gives the component impedance functions

$$Z_1 = \frac{R\omega_0(\omega_{-2}^2 - \omega^2)(\omega_{-a}^2 - \omega^2)(\omega_b^2 - \omega^2)(\omega_1^2 - \omega^2)}{j\omega(\omega_{-1}^2 - \omega^2)(\omega_{-b}^2 - \omega^2)(\omega_a^2 - \omega^2)(\omega_2^2 - \omega^2)} \tag{119}$$

and

$$Z_2 = \frac{R\omega_0(\omega_{-2}^2 - \omega^2)(\omega_{-b}^2 - \omega^2)(\omega_a^2 - \omega^2)}{j\omega(\omega_{-a}^2 - \omega^2)(\omega_b^2 - \omega^2)(\omega_2^2 - \omega^2)}. \tag{120}$$

These expressions, together with the cut-off frequencies

$$\left.\begin{aligned} \omega_1 &= 2\pi \cdot 1.1 \cdot 10^5, \\ \omega_{-1} &= 2\pi \cdot 0.9 \cdot 10^5, \end{aligned}\right\} \tag{121}$$

and the critical frequencies given by (114) and (117) completely solve our problem. The allocation of poles and zeros for the impedance functions (119) and (120) is given in Fig. 45. A

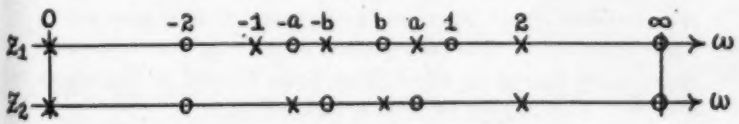


FIG. 45.

possible network representation for these impedance functions is shown in Fig. 46. The values of inductance and capacitance in these structures may be calculated from the above data by Foster's method.

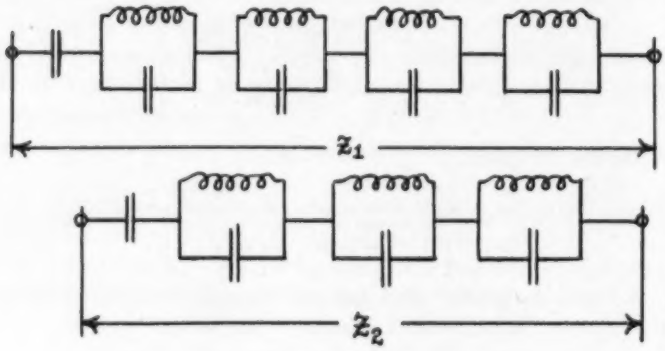


FIG. 46.

5. *Network Transformations.* It will probably have occurred to the reader that, although the Cauer method of design may be more logical and more flexible than the Zobel method, the lattice structure upon which it is based is very uneconomical as to the total number of coils and condensers involved. This is so because each component impedance enters into the structure twice. It is quite obvious therefore that the lattice structure per se cannot be used in any practical case. Fortunately, however, various other structures exist (an infinite number in fact) which are potentially equivalent to a lattice structure. Of these equivalent structures, there will be some that involve less coils and condensers than the original lattice. In any practical case it will therefore be necessary to determine one of these structures with fewer elements in terms of the component impedances Z_1 and Z_2 which have been derived in the above manner. This process we wish to describe now.

In Fig. 47 we have shown the usual arrangement for the lattice

structure. I_1 and I_2 are the input and output currents respectively and E_1 and E_2 are the corresponding voltages considered as voltage rises in the directions from the unprimed to the primed terminals. Under these conditions we can easily show

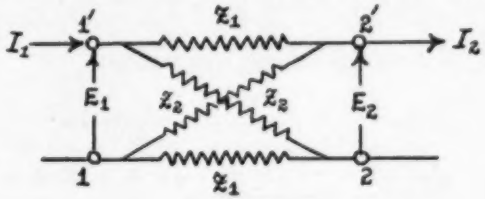


FIG. 47.

that the following relationships hold:

$$\left.\begin{aligned} E_1 &= \frac{Z_1 + Z_2}{2} I_1 + \frac{Z_1 - Z_2}{2} I_2, \\ -E_2 &= \frac{Z_1 - Z_2}{2} I_1 + \frac{Z_1 + Z_2}{2} I_2, \end{aligned}\right\} \tag{122}$$

and

$$\left.\begin{aligned} I_1 &= \frac{Y_2 + Y_1}{2} E_1 + \frac{Y_2 - Y_1}{2} E_2, \\ -I_2 &= \frac{Y_2 - Y_1}{2} E_1 + \frac{Y_2 + Y_1}{2} E_2. \end{aligned}\right\}^* \tag{123}$$

If in (122) we let $I_2 = 0$, then we obtain the open-circuit input and transfer impedances:

$$Z_{11} = \frac{E_1}{I_1} = \frac{Z_2 + Z_1}{2}; \qquad Z_{12} = \frac{E_2}{I_1} = \frac{Z_2 - Z_1}{2}. \tag{124}$$

If in (123) we let $E_2 = 0$, then we obtain the short-circuit input and transfer admittances:

$$Y_{11} = \frac{I_1}{E_1} = \frac{Y_1 + Y_2}{2}; \qquad Y_{12} = \frac{I_2}{E_1} = \frac{Y_1 - Y_2}{2}. \tag{125}$$

The two quantities in (124) or the two in (125) completely characterize the symmetrical lattice structure.

* The Y's are the reciprocals of the Z's.

If we consider $2 - 2'$ as the input and $1 - 1'$ as the output end, then the currents in (122) and (123) change sign, and we obtain for $I_1 = 0$, the open-circuit impedances:

$$Z_{22} = \frac{E_2}{I_2} = Z_{11}; \qquad Z_{21} = \frac{E_1}{I_2} = Z_{12}. \tag{124a}$$

and for $E_1 = 0$, the short-circuit admittances:

$$Y_{22} = \frac{I_2}{E_2} = Y_{11}; \qquad Y_{21} = \frac{I_1}{E_2} = Y_{12}. \tag{125a}$$

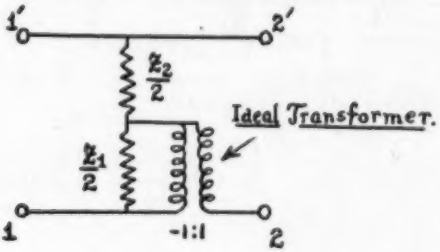


FIG. 48.

A network which has the open-circuit input and transfer impedances (124) and (124a), is shown in Fig. 48. This may easily be checked by inspection, bearing in mind that the ideal transformer is wound in the opposite sense on the primary and secondary. This is indicated in the figure by denoting the ratio of transformation with $-1:1$. The network of Fig. 48 is therefore equivalent to the symmetrical lattice structure. We see that the component impedances enter only once.

Another network, which has the short circuit input and transfer admittances (125) and (125a), is shown in Fig. 49. That this is so, may also be checked by inspection. This equivalent network also involves each component impedance only once.

The chief difficulty with the use of these equivalent circuits is due to the ideal transformers involved. An ideal transformer

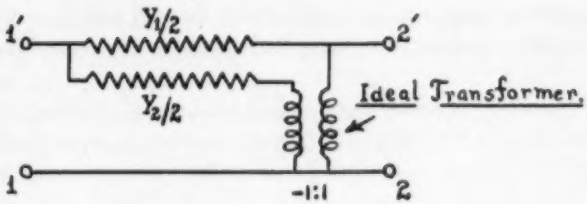


FIG. 49.

is one which has no losses, infinite impedance on open circuit, and zero impedance on short circuit, i.e. no leakage. Of these three requirements, the second is the most expensive to meet with a sufficient degree of approximation for most practical cases. How this second requirement may be eliminated by a modification of the structure of Fig. 48, with component impedances having the type of structure shown in Fig. 46, will now be shown.

It may be seen by inspection of the equations (122) or the open circuit input and transfer impedances (124), that an even more obvious equivalent structure than that of Fig. 48 is obtained by means of the T-structure of Fig. 50. This may, however,

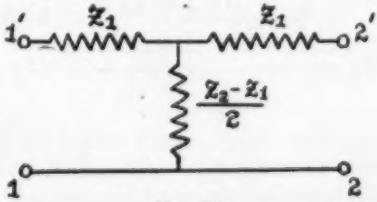


FIG. 50.

only be used when the impedance $(Z_2 - Z_1)$ can be realized physically. The most economical structure is in general found to be one which is based partly on the principle involved in Fig. 48 and partly on that of Fig. 50.

Suppose the component impedances are composed of two parts in series, as

$$\left.\begin{aligned} Z_1 &= \zeta_1 + \zeta_3, \\ Z_2 &= \zeta_2 + \zeta_4. \end{aligned}\right\} \tag{126}$$

Then we can apply the structure of Fig. 50 to ζ_1 and ζ_2, and that of Fig. 48 to ζ_3 and ζ_4 to obtain the realization shown in Fig. 51.

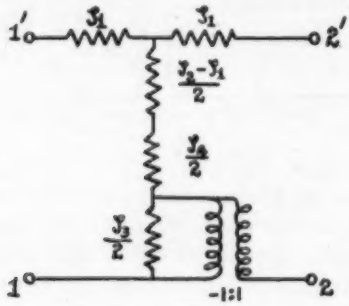


FIG. 51.

Now, in general, the component impedances may be determined in the form of Fig. 52. A possible realization according to

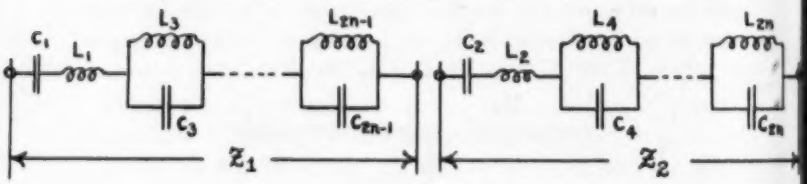


FIG. 52.

Fig. 51 is therefore shown in Fig. 53. This arrangement will generally contain the least number of elements. Instead of ideal transformers we merely need such with negligible loss and leakage. The inductances of these transformers on open circuit are $\frac{L_1}{2}, \frac{L_3}{2}, \cdots$, etc., and not infinite as required for the ideal transformer. This means that the coils $\frac{L_1}{2}, \frac{L_3}{2}, \cdots$, etc., are two identical windings in opposite sense on the same core.

One more point should be mentioned. In case $C_1 < C_2$, so that $C_1 - C_2$ would be negative it is only necessary to inter-

change Z_1 and Z_2. This, it will be remembered, has no effect either upon the characteristic impedance or attenuation functions.

6. *Conclusion.* Several matters may well be discussed in a little more detail before leaving this subject to the readers's own speculations. Whether or not the present method of design

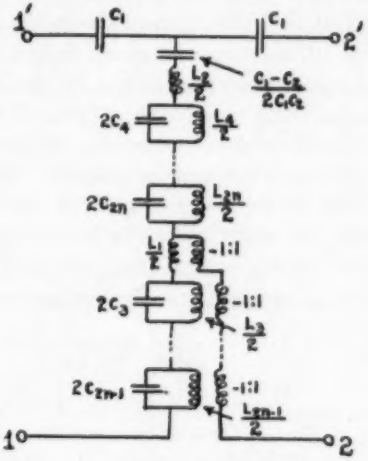


FIG. 53.

possesses any particular virtues from the practical standpoint is rather an open question. Economically, it will frequently lead to either identically the same structure (after transformation) as the Zobel method, or to one which involves about the same amount of material altogether as would be required by a Zobel design carried out so as to accomplish practically the same end. The advantages of the Cauer method, according to the writer's point of view, lie chiefly in its straightforwardness from a theoretical standpoint. In addition to this, it gives us a more complete picture of the situation than has been had up to the present time.

As already pointed out earlier, the Cauer method is not necessarily exclusive in that it enables one to prescribe the characteristic impedance and attenuation functions independently from the start. The Zobel method may be developed, by means of suitable charts, etc., to equal the Cauer method in this respect. However, it is certain that the latter will always have the advantage of being less involved.

In order to illustrate some of these things, consider again the low-pass filter which we discussed earlier in this paper, and for which the component impedances are illustrated by Fig. 2. Let us interchange these impedances, so that Z_1 becomes just a coil, and Z_2 a coil and condenser in series. According to eq. (12) both of these coils have the same inductance. Hence if we use the scheme of Fig. 50 for the realization of the final structure. we will find that the result takes the form shown in Fig. 54,

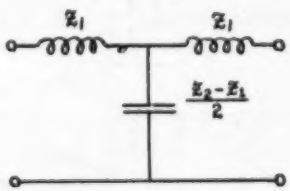


FIG. 54.

This, however, is a low-pass "constant-k" type of structure. This was to be expected since the functions $\sqrt{Z_1 Z_2}$ and $\sqrt{\frac{Z_1}{Z_2}}$ both have the general character of F_2. The attenuation is infinite at infinity because $\sqrt{\frac{Z_1}{Z_2}}$ is equal to unity at this point. It is not necessary that this should be so. The parameters of Z_1 and Z_2 may be chosen that the nominal value of $\sqrt{\frac{Z_1}{Z_2}}$ is changed to such a value that this function will equal unity at some finite point, which will become the point of infinite attenuation. This change may be made so as not to affect the char-

acteristic impedance in any way. For example, we have in general for this case:

$$\left.\begin{aligned} Z_1 &= jL_1\omega, \\ Z_2 &= jL_2\omega + \frac{1}{jC_2\omega} = \frac{1 - L_2C_2\omega^2}{jC_2\omega}, \end{aligned}\right\} \tag{127}$$

so that we get

$$\left.\begin{aligned} \sqrt{Z_1Z_2} &= \omega_1\sqrt{L_1L_2}\sqrt{1 - \left(\frac{\omega}{\omega_1}\right)^2}, \\ \sqrt{\frac{Z_1}{Z_2}} &= \sqrt{\frac{L_1}{L_2}} \cdot \frac{1}{\sqrt{1 - \left(\frac{\omega_1}{\omega}\right)^2}}, \end{aligned}\right\} \tag{128}$$

where

$$\omega_1 = \frac{1}{\sqrt{L_2C_2}} \tag{129}$$

is the resonant frequency of Z_2 and hence the cut-off.

Instead of making $L_1 = L_2$, i.e. making the nominal value of $\sqrt{\frac{Z_1}{Z_2}}$ equal to unity, we may let

$$\frac{L_1}{L_2} = K^2. \tag{130}$$

Then

$$\left.\begin{aligned} \sqrt{Z_1Z_2} &= \frac{L_1\omega_1}{K}\sqrt{1 - \left(\frac{\omega}{\omega_1}\right)^2}, \\ \sqrt{\frac{Z_1}{Z_2}} &= \frac{K}{\sqrt{1 - \left(\frac{\omega_1}{\omega}\right)^2}}, \end{aligned}\right\} \tag{131}$$

so that the nominal characteristic impedance is

$$R = \frac{L_1\omega_1}{K}. \tag{132}$$

From the second relation (131) we see that the point of infinite attenuation falls where

$$K = \sqrt{1 - \left(\frac{\omega_1}{\omega}\right)^2},$$

or at:

$$\omega = \frac{\omega_1}{\sqrt{1 - K^2}}. \tag{133}$$

The parameter K is evidently restricted to

$$0 < K < 1. \tag{134}$$

The structure is now changed from that shown in Fig. 54 to the one given by Fig. 55. This we recognize as a Zobel mid-series equivalent single M-derived type, which by his method is derived so that its characteristic impedance equals that of the "constant-k" type. When obtained from the Cauer standpoint, however,

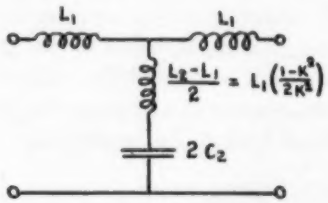


FIG. 55.

no particular effort was required to keep the characteristic impedance invariant. This takes care of itself since it is independent of the attenuation function anyway. Shifting the point of infinite attenuation merely required a slight change in the nominal value of $\sqrt{\frac{Z_1}{Z_2}}$.

In presenting the material contained in this paper, the writer's object has not been so much to give the reader a finished tool for the design of filter structures, but rather to acquaint him with the general ideas involved in this new process, and to give him a few landmarks as to how it is related to the Zobel method with which he is already familiar. There are still a number of matters which need to be worked out in connection with the Cauer method. For example, for the B.P. and B.E. types the

method as outlined in this paper can handle only such cases for which the band width is small compared to the distance from the center of the band to the origin. The method must be further developed to include cases for which this condition is not met to a sufficient degree of approximation.* It can also be extended to handle cases involving more than one pass or attenuation band, but this still remains to be done.

It is also necessary to show how to connect the Cauer type of filters in series or parallel on their input sides for the purpose of separating communication channels coming over a single line. For the Campbell type, Zobel has solved this problem for the case of complementary filters by means of his "x-terminations," and Bode [5] has generalized this point of view to include any other classes. Undoubtedly, Bode's method could be applied to the Cauer filter as well, but not with very good overall economy.

The Cauer method very materially broadens our general understanding of this subject; but a method which attacks the problem, not only of filters but of corrective networks generally, in terms of the desired characteristics only, without reference to any particular basic structure, and arrives at a most economical structure sheerly by reason of its own completeness, still remains to be developed.

* Cauer in his publication (see footnote 1) has given certain substitutions which are intended to cover such cases, but the form of his results is so complicated that they are of little practical value.

[5] H. W. Bode, "A Method of Impedance Correction," *B. S. T. J.*, Oct. 1930.

CORRECTIONS TO THE PAPER "THE PROBLEM OF PLATEAU FOR TWO CONTOURS"

By Jesse Douglas

I have noticed the following errata in my paper "The problem of Plateau for two contours," this Journal, vol. X, no. 4 (Dec. 1931). None of them affects any result or proof.

1. Page 325, formula (2.9). Replace w in the denominator by w_0, making the formula read

$$G(x, y; u, v) = \Re \log \frac{S(z, w)}{S_1(z, w_0)}.$$

2. Page 325, ninth line from bottom. Read

$$\frac{\partial M}{\partial w} = \frac{\partial (iN)}{\partial w}, \qquad \frac{\partial M}{\partial w_0} = -\frac{\partial (iN)}{\partial w_0}.$$

3. Page 327, seventh line. Instead of (2.14) read (2.15).
4. Page 329, formula (3.2). Read

$$\log S(z, w) = \log (z - w).$$

5. Page 332, second line from bottom. Read (3.14_1) instead of (3.14).
6. Page 336, formula (4.3). For exponent of z in the denominator, read

$$\frac{\log w_0}{2 \log q}.$$

7. Page 339, formulas (5.4) and (5.6). Omit the coefficient 4 from the term containing the reciprocal of $-\log q$.
   The same correction in the tenth line from the bottom of page 339, and in the first formula on page 359.
8. Pages 339, 340, formulas (5.7), (5.8). Delete the term $3\pi\alpha^2$ from the numerator.
   The term $3\pi'\alpha^2$ to be deleted in the fifth line on page 357.

THE FIELD DUE TO TWO EQUALLY CHARGED PARALLEL CONDUCTING CYLINDERS

By Hillel Poritsky *

The case of two parallel cylinders with equal charges of opposite sign is a familiar one. As is well known, the field may be expressed in terms of logarithmic functions, and yields circles both for the lines of force and the equipotentials. On the other hand, it will be shown that the case of equal charges of the *same* sign requires doubly periodic or elliptic functions for its treatment. We shall confine ourselves for the present to the case of cylinders of the *same* radius a.

Let Q be the charge on each cylinder per unit length of axis, and $2b$ the (shortest) distance between the cylindrical surfaces, In a plane perpendicular to the cylinder axes and cutting the cylindrical surfaces in circles C_1, C_2 place a coördinate system with the x-axis intersecting the cylinder axes and the origin half way between the (centers of the) circles C_1, C_2. The problem resolves itself into finding a harmonic function u (the potential) which takes on the same constant value along the circle C_1, C_2, and whose conjugate harmonic function v (or flux function) increases by $4\pi Q$ for a path which encircles either C_1 or C_2 alone in the positive sense; moreover, at infinity u is to become infinite like $2Q \log (x^2 + y^2)$. This problem is treated best by means of complex quantities and the theory of analytic functions.

In the x, y-plane we introduce the complex variable:

$$z = x + iy, \qquad i^2 = -1.$$

As is known the potential u and its conjugate harmonic function v define an analytic function:

$$w = w(z) = u + iv. \tag{1}$$

* Dr. Poritsky is connected with the General Electric Company at Schenectady, New York.—Ed.

. First we shall carry the problem from the z-plane to a z_1-plane, where

$$z_1 = x_1 + iy_1 = \log\frac{c+z}{c-z}, \qquad c = \sqrt{b(b+2a)}. \tag{2}$$

This carries the circles C_1, C_2 into the straight lines:

$$x_1 = \pm d, \tag{3}$$

where

$$d = \log\frac{\sqrt{b+2a}+\sqrt{b}}{\sqrt{b+2a}-\sqrt{b}}. \tag{4}$$

The region outside the circles is transformed into the region between the lines (3) and the original field goes into a field which is periodic in the y_1-direction of period 2π. Since the potential u is constant along $x_1 = \pm d$, the field may be extended across each of these boundaries to a width $2d$ on each side, then extended across the new boundaries, etc., resulting finally in a doubly periodic field the periods being $4d$ and $2\pi i$. This field consists of sinks of lines of force at $z_1 = \pi i$, the image of $z = \infty$, and of its congruent points, and of sources at $z_1 = 2d + \pi i$ and its congruent points. At each source $8\pi Q$ lines start and on equal number end at each sink. It is now obvious why doubly periodic or elliptic functions are needed for the description of the field.

To pave the way for the latter, consider the function,

$$W = e^{w/4Q} \tag{5}$$

in the z_1-plane. At infinity in the z-plane u becomes infinite like $4Q \log |x + iy|$. Therefore at infinity in the z-plane $w - 4Q \log z$ is analytic in $1/z$, and hence W has a simple pole at $z = \infty$, and consequently at all the sinks in the z_1-plane. Similarly at each source W has a simple root. W is therefore meromorphic in z_1. It is moreover periodic in z_1 of period $4d$ and changes sign for an increase of $2\pi i$ in z_1—the latter is due to the fact that $w/4Q$ increases by πi for a path which encircles

one of the cylinders. Utilizing these properties of W we shall now obtain an explicit expression for it. First however we shall introduce the auxiliary transformation,

$$z_2 = (z_1 - \pi i)/4d, \tag{6}$$

which converts the periods $4d$, $2\pi i$ to unity and ω given by

$$\omega = 2\pi i/4d = \pi i/2d, \tag{7}$$

and shifts the origin to a pole of W. The poles of W are now at

$$z_2 = m + n\omega,$$

while the roots are at

$$z_2 = m + \tfrac{1}{2} + n\omega,$$

where m, n are integers. W also satisfies the functional equations:

$$\begin{aligned} W(z_2 + 1) &= W(z_2), \\ W(z_2 + \omega) &= -W(z_2). \end{aligned} \tag{6}$$

Consider now the function,

$$f(z_2) = \vartheta_1(z_2 - 1/2)/\vartheta_1(z_2),$$

where ϑ_1 is defined by

$$\vartheta_1(v) = 2\sum_{n=0}^{\infty}(-1)^n q^{[n+(1/2)]^2} \sin(2n+1)\pi v, \qquad q = e^{\pi i\omega}.$$

Recalling the following properties of ϑ_1:

1. $\vartheta_1(v)$ integral or entire in v,
2. $\vartheta_1(v)$ has simple roots at $v = m + n\omega$ and nowhere else,
3. ϑ_1 satisfies the functional equations:

$$\begin{aligned} \vartheta_1(v + 1) &= -\vartheta_1(v), \\ \vartheta_1(v + \omega) &= -q^{-1}e^{-2\pi i v}\vartheta_1(v),^1 \end{aligned}$$

one proves that $f(z_2)$ has the same zeros and poles as W, and satisfies the functional equation (6). Hence $W(z_2)/f(z_2)$ is

[1] See for instance, Pierpont, *Functions of a Complex Variable*, Chap. XII, particularly pp. 428, 429.

elliptic of periods 1 and ω and has no zeros or poles. By a well known theorem it must therefore reduce to a constant. Hence

$$w = 4Q \log f(z_2) + \text{constant}.$$

We have thus obtained an explicit expression for the field.

To find the field intensity in the z-plane we need the values of dw/dz. This is found from

$$\frac{dw}{dz} = \frac{dw}{dz_2}\frac{dz_2}{dz_1}\frac{dz_1}{dz}.$$

Carrying out the differentiations at the point of maximum field intensity $z = b + 2a$ corresponding to $z_2 = 1/4$, we obtain for the maximum field intensity $|dw/dz|$ there:

$$\frac{2\pi Q}{ad}\frac{\sqrt{b}}{\sqrt{b+2a}}\frac{1 + 3q^{1\cdot 2} - 5q^{3\cdot 2} - 7q^{3\cdot 4} + 9q^{4\cdot 5} + \cdots}{1 - q^{1\cdot 2} - q^{3\cdot 2} + q^{3\cdot 4} + q^{4\cdot 5} - \cdots}.$$

In addition to the initially given quantities Q, a, b this expression involves q and d, given by

$$q = e^{-\pi^2/2d}, \qquad d = \log\left[(\sqrt{b+2a} + \sqrt{b})/(\sqrt{b+2a} - \sqrt{b})\right].$$

The case in which the two cylinders touch each other may be obtained as a limiting case from the above by letting b approach zero, but it is instructive to solve it directly. Now the region outside C_1 and C_2 is simply connected and there is no possibility of going around C_1 or C_2 separately. As a result, when the region outside these circles is plotted on a region between two parallel lines, the field is no longer periodic in the direction of the lines. After reflection across the lines a simply periodic field results which may be described by means of circular functions. Thus

$$z_1 = 1/z$$

takes the place of equation (2) and transforms the region outside C_1, C_2 into the region between $x_1 = \pm 1/2a$. The field which results after reflection in these lines and further reflection has

sinks at $z_1 = 2n/a$ and sources half-way between the sinks. As above we now show that $W = e^{w/4Q}$ is periodic of period $2/a$, that it has poles and zeros at these sinks and sources respectively, and in a similar manner prove that it is equal to a constant multiple of cot $(\pi a z_1/2)$. The maximum field intensity now turns out to be $Q\pi/a$.

While we have confined ourselves above to the case of cylinders of equal radii, it is evident that cylinders of unequal radii but with the same charge on each one can be handled in a similar way by making proper changes in the conformal transformation (2). Again, the case of unequal charges may be treated by resolving the potential into a sum of one due to equal like charges, and one due to equal and opposite charges.

Since the above was written, the author has run across an article by F. J. W. Whipple in the Proceedings of the Royal Society, A, vol. 96, (1920), p. 465, dealing with the same subject and entitled "Equal Parallel Cylindrical Conductors in Electrical Problems." There is enough difference in the two treatments to justify the publication of this article which is in many ways more elementary and lends itself easier to applications. The author's interest in the problem was aroused by questions of field intensity and breakdown danger; Mr. Wipple's approach is thru skin effect problems.

RITZ'S ELECTRODYNAMIC THEORY

By William Hovgaard

I. Introduction

Walther Ritz does not need any introduction to the scientific world. His brilliant work in theoretical and experimental physics as well as in mathematics is well known, but although Ritz was a physicist as well as a mathematician, he was first of all a philosopher whose chief interest was in abstract speculations on the principles and fundamentals of science. It was in this capacity that he took up the subject of electrodynamics, attempting to prepare a synthesis of a new theory of this science. Unfortunately he had to leave this work uncompleted.

Ritz was born in Sion, Switzerland, in 1878. His short span of working years extended from 1900 to 1908, but already in 1904 his health began to fail due to chronic pleurisy, which henceforth interfered seriously with his work. It was at that time that he was drawn into a study of the science of electrodynamics, which was in a state of rapid development under the influence of the theories of Lorentz, Poincaré, Einstein, and others, and this subject occupied the greater part of his time during his last years. In 1908 he published his first and most comprehensive memoir on electrodynamics in *Annales de Chimie et de Physique*, in which he subjected Maxwell's theory, as modified by Lorentz, to a critical analysis and outlined a new theory of electrodynamics. This memoir was followed by several papers on the same and allied subjects such as the theory of the ether and of gravitation.

Ritz continued his work on the new theory till the last. His ardor and enthusiasm for science were at once the strongest reason for his desire to live, but also, according to his physician, the principal obstacle to his recovery. His failing health and

early death prevented the completion of the new theory. He died in July 1909 in his thirty-first year. The day of his death he said to his sister who nursed him: "Take good care of me, sister. I do so want to live some years more for the sake of science." Had this wish been fulfilled, he would probably have given to the world a new and fruitful theory, which might have profoundly changed the development of this branch of science.

His first memoir: "*Recherches Critiques sur l'Électrodynamique Générale*," [1] hereafter referred to as reference (A) is divided into two parts, the first of which contains a critical study of the Maxwell-Lorentz electrodynamic theory, the second, an outline of a new theory closely associated with the emission theory of light. Later in 1908 he published in *Archives de Genève* [2] a paper, which is practically an abstract of the first part of his first memoir, and is well adapted to serve as an introduction to his new theory. A complete translation of this paper is here given in Chapter II, and is followed in Chapter III by a brief abstract of the new theory. Chapter IV contains some concluding remarks by the author of this paper.

Ritz's criticism of Lorentz's theory centered in the concept of the ether where most of the difficulties in that theory originate. Lorentz regarded the ether as immovable and was thus forced to adopt an absolute system of coordinates, independent of the movements of matter. The earth was conceived as moving through the ether without disturbing it and it followed that there must be an "ether-drift," which it should be possible to detect. But this proved impossible. Even the extremely refined measurements of the Michelson-Morley experiments failed to give any indication of an ether-drift and it became necessary to resort to new hypotheses. This was the parting of the ways. While Ritz, as explained below, proposed to adopt

[1] *Annales de Chimie et de Physique*, 8 Series, Vol. XIII, Febr. 1908, pp. 145–275.

[2] 4th. Period, vol. 26, Sept. 1908: *Recherches Critiques sur les Théories Électrodynamique de Cl. Maxwell and H.-A. Lorentz.*

an emission theory associated with periodic motion, and the complete abandonment of the ether, Lorentz and Fitzgerald advanced the so-called "contraction hypothesis" according to which all bodies should contract when they move through the ether. Another theory was advanced by Einstein, who interpreted the Michelson-Morley experiments as showing simply that the velocity of the propagation of light was the same for an observer in motion as for one at rest, and on the basis of this postulate he constructed his *Special or Restricted Theory of Relativity*. Lorentz accepted Einstein's theory, and attempted at the same time to uphold the theory of the ether, but this led to highly artificial and complicated conceptions both in electrodynamics and in optics. Besides the contraction hypothesis, it was necessary that every observer should carry a complete ether about with him,[3] since otherwise the velocity of light relative to him would not be constant. The ether, although immovable relative to an observer, must still be supposed to be endowed with motion and momentum and subject to stress and strain in order to account for the magnetic energy and the electromagnetic forces. The increasing complexity of the properties which had to be ascribed to the ether in order to account at the same time for all phenomena, gradually forced it into the background, although many physicists still maintain that it is a necessary concept. On the other hand, if the Theory of Relativity is accepted, all phenomena can be accounted for whether an ether exists or not. It is significant that the word "ether" does not appear either in the *Contents* or the *Index* of Eddington's *Mathematical Theory of Relativity*. In Jeans' work on *Electricity and Magnetism*[4] on the other hand, the whole electromagnetic theory is based on the existence of an ether, but in a final chapter on Relativity it is definitely discredited.

In fact the concept of the ether and of an absolute system of coordinates is incompatible with Einstein's theory and this latter, which requires only the acceptance of one simply stated

[3] Jeans, *Electricity and Magnetism*, 5th Ed., p. 619.

[4] Fifth Edition, pp. 618–621.

postulate, was soon widely adopted. The postulate can be formulated concretely thus, that to any observer the velocity of light is always 300,000 km. per sec. whatever be the motion of the source of light or of the observer. If the velocity of a ray of light is c and that of one observer $+v$ and of another $-v$, moving respectively in the same and opposite directions as the ray of light, we shall have:

$$c + v = c - v = c,$$

where v can have any value smaller than c. This contradiction is not, of course, accessible to understanding when viewed in the light of Newtonian mechanics, but can be expressed mathematically. This was done by the so-called Lorentz-Transformation, which establishes certain relations between time and spatial coordinates, but which does not remove the contradiction. By these relations a new significance is given to space and time, which are no longer independent of one another; the concepts of absolute time and simultaneity lose their meaning; length and time scale differ from one moving body to another and the classical principles of kinematics and dynamics have to be abandoned; the rule of the parallelogram of velocities is to be considered as a first approximation, valid at small velocities only, and it is necessary to abolish the notion of rigid bodies and the invariability of ponderable mass. Ritz said, "it would be deplorable, for our economy of thought, if we had to accept such complications." [5]

The difficulty or impossibility of conceiving a relation between space and time was apparently relieved by the genius of Minkowski, who devised a four-dimensional picture involving time and the three spatial coordinates, but the time coordinate represents an imaginary quantity, $ct\sqrt{-1}$. This picture, of course, was purely mathematical and descriptive. It simplified the mathematical presentation and, by its analogy to the three-dimensional system, it seemed also to facilitate the conception of the theory,

[5] Reference A, p. 148: "Il serait regrettable, pour l'économie de notre pensée, qu'il fallut admettre de telles complications."

diverting, as it did, the attention from the inherent unreality of space-time.

In the so-called *General Theory of Relativity* we meet such terms as "flat space-time," "curved space-time," etc., referring of course to four-dimensional space. These terms need only be mentioned in order to indicate how far the theory has diverged from Newtonian concepts of the physical world. The analogies, in which articles and works on Relativity abound, cannot explain the inconceivable.

The reason why this theory was so successful in spite of the extraordinary demands which it made on imagination, was that it was confirmed in a remarkable manner by a number of experiments and observations; in fact, it even suggested or predicted new physical phenomena, which were afterwards proved to exist. The superior mathematical ability of its sponsors guaranteed the soundness of the mathematical deductions.

So great has become the prestige of the Theory of Relativity that physical tests and observations are often considered subservient to it, instead of as independent building materials in an inductive process of reasoning. In other words, the process, outlined above, is regarded as reversible, which means the adoption of a new hypothesis. The Lorentz-Transformation, which was at first a purely mathematical operation, has acquired the status of a law of nature. According to Einstein [6] the Theory of Relativity prescribes for all general natural laws that they shall be covariant relative to Lorentz's Transformations. Eddington says:[7] "Our present exposition is in the main deductive. We start with a general theory of world-structure and work down to the experimental consequences, so that our progress is from the general to the special laws, instead of *vice versa*."

Yet, the Theory, in spite of all its successes, is essentially a mathematical structure, probably unassailable as such, but based

[6] "Über die Spezielle und die Allgemeine Relativitätstheorie," A. Einstein, § 14.

[7] *The Mathematical Theory of Relativity*, A. S. Eddington, 1930, p. 41.

primarily on one physical unexplained fact which may perhaps in the future be differently interpreted and physically explained.

It is to be noted also that the agreement which has been found between the Theory and observation depends in most cases on second order terms, that is, on extremely small quantities, difficult to observe and, as shown by Ritz, it is quite possible to account for all or most of those phenomena without departing from Newtonian mechanics.

The Einstein theory did not, of course, escape criticism, but the only, or at least the most serious attempt to find another solution was that made by Ritz. As mentioned above, Ritz commenced his study with a critical analysis of the Maxwell-Lorentz theory, published in the First Part of his great memoir. We shall here give a brief abstract of this analysis, supplementing the translated paper here given in Chapter II.

Ritz accepted the theory of the electron, but rejected the ether; he maintained that from a strictly logical point of view, Maxwell's electric and magnetic forces, which apparently play such an important part in the theory, can be entirely eliminated from it. In reality Maxwell's theory in the last analysis deals only with certain relations between space and time, and thus we can return to the old elementary actions with the sole difference that they are no longer instantaneous. The field equations admit of an infinite number of solutions, that are incompatible with experience; in order to remove these solutions it is necessary to adopt the retarded potentials, which actually constitute a new hypothesis, but then the field equations become useless. In order to express this hypothesis clearly, it is necessary to give prominence to the elementary actions, and therewith to renounce the fundamental idea of Maxwell, who rejected them. The magnitude and direction of the field can be obtained only by placing a body in it and by observing the mechanical forces to which it is subject. We have really no direct knowledge about anything except the mechanical force F, and that only at points where there is electrified matter. Thus it is sufficient to know the formula which gives F as the

sum of the elementary actions exerted by one element of charge on another. This mode of representation is, as regards the facts, exactly equivalent to the method that starts from the field equations, which serve only as a purely mathematical intermediary. We may, if we please, do entirely without the notions of electric and magnetic fields.

Ritz found that Poynting's theorem expresses the law of energy only when the fields are replaced by their expressions derived from the retarded potentials.

The notion of electromagnetic fields is inapplicable to gravity. The condition for stability in a continuum, whether elastic or not, is that the energy shall be a minimum when the deformation is zero; in this case it is a maximum when the force is equal to zero. Hence, the gravitational field will be in unstable equilibrium at infinity and whenever the gravitational force is zero.

Since the ether acts on the ions without itself suffering any action, Newton's principle of action and reaction is not satisfied by Lorentz's theory. Two electrons in uniform translatory motion are subject to forces, which should form a couple if the electrons were rigidly connected, but this has been disproved experimentally (Trouton and Noble). As far as the velocity terms in the elementary force are concerned, this inequality in action and reaction constitutes a grave objection to Lorentz's theory.

The calculation of the electrodynamic mass rests entirely on the consideration of the movement of the electrified body relative to the ether; hence, it is the *absolute velocity* that enters into the formula. The interpretation put upon Kaufmann's experiments on the deflection of β-rays, must yet be considered doubtful in spite of their accord with the theory of Abraham.

Ritz concludes the First Part of his memoir by saying that we must abandon the picture of the ether. The movement of light is relative like all others, only relative velocities play a part in the laws of nature. We must abandon the partial differential equations and the notion of the field in the measure

in which it introduces absolute motion. This conclusion, he says, being purely negative needs two complements: a simple picture for a new mode of movement of light and a proof that a satisfactory theory is possible. This forms the object of the Second Part of his memoir, of which a brief abstract is given in Chapter III.

We shall first in Chapter II give a complete translation of the aforementioned paper published in *Archives de Genève*. The text is followed as closely as possible, and all additions or comments by the translator are placed in footnotes preceded by the letters *Tr.*

II. Translation of Ritz's Paper in Archives de Genève "A Critical Investigation of Maxwell's and Lorentz's Electrodynamic Theories"

The history of the new ideas introduced by Maxwell in the science of electricity and the theories which are derived from them, form decidedly one of the most interesting chapters in the history of science, especially from a psychological point of view. It is well known how the scientific world—accustomed to the transparent clarity which gave such high esthetic value to the classical theories of mathematical physics—revolted against these new ideas, which disturbed the established order, and which, it must be admitted, seemed at first strangely confused. The first publication of Maxwell dates from 1856: "On Faraday's Lines of Force"; but even thirty years later it required all the authority of Helmholtz to obtain a hearing for the new theory, so that even if it were not accepted, it could at least be deemed worthy of some interest. It was the experiments of Hertz and his followers, which, by demonstrating the identity of light and electric oscillations, and by confirming thus the brilliant concepts of Maxwell, broke down the last opposition to his theory and gave it the citizenship in physics. It was afterwards recognized that the obscurities in Maxwell's theory had their origin largely in the fact that he combined two very different lines of thought;

one, a tentative explanation of the electrical actions by means of the properties of the medium through which they are transmitted (an explanation which led Maxwell to various accessory hypotheses, and where, in spite of all efforts, he failed completely as far as electrostatics is concerned); another, a purely phenomenological description employing partial differential equations and a hypothesis as to the electromagnetic energy, in which appear certain vectors that characterize the electric and magnetic state of the body. By choosing the second method, we shall avoid many difficulties.

Maxwell's theory, extended by Hertz to bodies in motion, does not agree with certain optical experiments (aberration, Fizeau's experiments, etc.), nor with those of Eichenwald on the action of movable dielectrics. The new form which Lorentz gave to Maxwell's theory, on the other hand, is in perfect agreement with those experiments. He moreover, simplified the equations considerably by re-adopting the hypothesis of Fechner and Weber, that all electric currents are currents of convection, that is, that they are due to a movement of electricity, a hypothesis which is confirmed more and more by recent research. The atomic constitution, with which Lorentz endowed electricity, gives a clearer and more precise view of the phenomena. Finally, by considering the ether as immovable and as existing even in the interior of the atoms, he suppressed an indeterminacy in Maxwell's theory, which it had not been possible to correct previously. This indeterminacy resulted from the movements of the ether, which were also required according to the theory of Hertz, although not clearly specified, but which had never been proved to exist experimentally. Finally, the mutual and complete compenetrability of ether and matter [8] explained how bodies can move through the ether without meeting any resistance, and why the "ether drift" of 30 km. per second, which, according to Fresnel and Lorentz, flows through the earth in its movement around the sun, has never been detected, even by the most delicate observations.

[8] Tr. (assumed by Lorentz).

Reducing in this way Maxwell's theory to its simplest form and avoiding many mathematical difficulties, Lorentz filled in the huge gap which separated it from the classical theories founded on action at a distance, and he defined clearly the mutual relations that exist between the equations of Weber and Clausius on the one hand, and those of Maxwell and his own on the other hand. As will be shown below, these relations are much closer than would appear at first sight.

But the theory thus simplified presents the further advantage that it permits a more rigorous criticism of the principles on which it is based. These principles are of different kinds. There are, first, the experimental bases of the theory: could it be that the experiments, which at first sight seem to confirm the theory so amply, have unconsciously been directed toward certain points, leaving others, equally important, in the background? In other words, what are the modifications which could be made in Lorentz's formulas without coming into conflict with any experimental facts?

Second, what is the true significance of the vectors called the electric force E and the magnetic force H, which enter into the field equations? And how do we pass from these to the experimental facts which they are to represent? Analogous questions have been made in mechanics, as well known, and have not been answered until quite recently. Now, through the concept of electromagnetic mass, and due to the impotence of the theory to explain the phenomena by means of the mechanical properties of the ether, modern physicists have been led to conceive inversely an electromagnetic origin of the laws of mechanics. Thus electrodynamics has become the pivot for a new conception of nature, which takes the place of the old mechanics. It is therefore of particular importance that no cloud shall obscure the logical foundations of this vast intellectual structure.

Among these foundations is the hypothesis of an absolute system of coordinates, but the experiments of Michelson and

Morley, as well as others of more recent date, have on this point given a formal negation of this theory. As in mechanics, a uniform translatory motion of a system does not seem to have any influence on the optical and electromagnetic phenomena. Lorentz, Einstein, Poincaré and others have therefore inquired into what new hypotheses must be introduced in order to reconcile this fact with the theory, *without touching the fundamental field equations*. The result is that it has been found necessary to abandon the classical concept of universal time, to make of simultaneity a quite relative notion, to suppress the concept of the invariability of mass as well as that of rigid bodies, to abandon the axioms of kinematics and the parallelogram of velocities, etc. When a grain of radium emits β-rays in two opposite directions at a velocity of 250,000 km. per sec., the relative velocity of the two rays is not 500,000 km. per sec. but is equal to 294,000 km. per sec. Further, two times which are equal for an observer A, or two simultaneous events, are no longer equal or simultaneous for an observer B, who is moving relative to A. It is curious and worth noting that a few years ago it would have been thought sufficient, in order to refute a theory, to show that it entailed only one or other of the consequences here enumerated, but now the equations of Maxwell are considered so absolutely *inviolable*, that these consequences do not deter anyone. Rather than to modify these equations more or less profoundly, it has been decided to sacrifice kinematics, the notion of time, etc. After having ignored more or less systematically a fruitful theory for thirty years, we fall into the opposite extreme. Do these equations really deserve such extreme confidence?

The answer to this question is decidedly negative, and I shall here present a résumé of the criticisms to which Maxwell-Lorentz theory gives rise and of the experimental uncertainties which it entails. In another memoir [9] the reader will find the proofs of the formulas in detail.

Let us first briefly review Lorentz's fundamental equations.

[9] Reference A.

Let H be the magnetic vector, E the electric vector (or force), c the velocity of light, $\rho(x, y, z, t)$ the electric density at a point x, y, z, at the time t, v the velocity of this charge associated with an ion or electron. Finally, denote by A_l the component of a vector A in the direction l. Then we have:

$$\left.\begin{aligned} \frac{dH_z}{dy} - \frac{dH_y}{dz} &= \frac{1}{c}\frac{dE_x}{dt} + 4\pi\rho\frac{v_x}{c} \\ \frac{dH_x}{dz} - \frac{dH_z}{dx} &= \frac{1}{c}\frac{dE_y}{dt} + 4\pi\rho\frac{v_y}{c} \\ &\cdots\cdots\cdots \end{aligned}\right\} \tag{I}$$

$$\left.\begin{aligned} \frac{dE_z}{dy} - \frac{dE_y}{dz} &= -\frac{1}{c}\frac{dH_x}{dt} \\ \frac{dE_x}{dz} - \frac{dE_z}{dx} &= -\frac{1}{c}\frac{dH_y}{dt} \\ &\cdots\cdots\cdots \end{aligned}\right\} \tag{II}$$

$$\frac{dE_x}{dx} + \frac{dE_y}{dy} + \frac{dE_z}{dz} = 4\pi\rho \tag{III}$$

$$\frac{dH_x}{dx} + \frac{dH_y}{dy} + \frac{dH_z}{dz} = 0 \tag{IV}$$

$$\frac{d\rho}{dt} + \frac{d(\rho v_x)}{dx} + \frac{d(\rho v_y)}{dy} + \frac{d(\rho v_z)}{dz} = 0. \tag{V}$$

The field so created in the ether exerts on an element of charge $\rho d\tau$ a mechanical force represented by the vector $F\rho d\tau$ where:

$$\begin{aligned} F_x &= E_x + \frac{1}{c}(v_y H_z - v_z H_y); \\ F_y &= E_y + \frac{1}{c}(v_z H_x - v_x H_z); \qquad \cdots. \end{aligned} \tag{VI}$$

Lorentz considers only a certain particular integral of the system (I) to (V), obtained by means of the "retarded po-

tentials." It is shown, in fact, that all integrals of this system can be written in the form:

$$E_x = -\frac{d\Phi}{dx} - \frac{1}{c}\frac{dA_x}{dt}; \qquad E_y = -\frac{d\Phi}{dy} - \frac{1}{c}\frac{dA_y}{dt}; \qquad \cdots \quad \text{(VII)}$$

$$H_x = \frac{dA_z}{dy} - \frac{dA_y}{dz}; \qquad H_y = \frac{dA_x}{dz} - \frac{dA_z}{dx}; \qquad \cdots. \quad \text{(VIII)}$$

The functions Φ (a scalar potential) and A_x, A_y, A_z (components of the vector potential) with their first derivatives are continuous throughout all space, vanish at infinity and satisfy the equations:

$$\frac{1}{c^2}\frac{d^2\Phi}{dt^2} - \Delta\Phi = 4\pi\rho \qquad \text{(IX)}$$

$$\frac{1}{c^2}\frac{d^2A_x}{dt^2} - \Delta A_x = \frac{4\pi\rho v_x}{c}; \quad \frac{1}{c^2}\frac{d^2A_y}{dt^2} - \Delta A_y = \frac{4\pi\rho v_y}{c}; \quad \cdots \quad \text{(X)}$$

$$\frac{dA_x}{dx} + \frac{dA_y}{dy} + \frac{dA_z}{dz} = -\frac{1}{c}\frac{d\Phi}{dt} \cdots. \qquad \text{(XI)}$$

The functions:

$$\Phi(x, y, z, t) = \int\int\int \frac{[\rho']d\tau'}{r} \qquad \text{(XII)}$$

$$A_x = \frac{1}{c}\int\int\int \frac{[\rho' v_x']}{r} d\tau';$$

$$A_y = \frac{1}{c}\int\int\int \frac{[\rho' v_y']}{r} d\tau'; \qquad \cdots, \qquad \text{(XIII)}$$

are *particular* integrals of the system IX, X, XI; they have the same form as Newtonian potentials, with the difference that the value of ρ, instead of being taken at the point $x'y'z'$ at the time t, is taken at the *earlier* time $t - \frac{r}{c}$ where r is the distance between the points xyz and $x'y'z'$. Following Lorentz, this is indicated by the notation $[\rho']$, $[\rho' v_x']$, etc., or, more generally:

$$[f] = f\left(x', y', z', t - \frac{r}{c}\right).$$

The field is now completely determined, and on introducing these values of Φ and A in VII, VIII and VI we obtain an analogous expression, that is, a triple integral extended to the "retarded forces"; a rather complicated expression which we will not write down,[10] but which expresses the force to which an electric point of unit charge is subject. This force is produced by elementary actions analogous to those considered in the old electrodynamics, and found also in Gauss, except that the transmission is not instantaneous. In the case where two electrons are at a finite distance apart and under certain general conditions, which are here of no importance, we obtain from that formula the following expression [11] for the force exerted by an electron e' of velocity v_x', v_y', v_z' and acceleration w_x', w_y', w_z' on an electron e which has a velocity v:

$$F_x = ee' \left\{ K_x + \frac{1}{c}\left[(v_xK_x + v_yK_y + v_zK_z)\cos(rx) - v_rK_x\right] \right\}; \quad \cdots, \qquad \text{(XIV)}$$

where K is the electric force at x, y, z, given by

$$K_x = -\frac{w_x'}{c^2 r\left(1 - \frac{v_r'}{c}\right)^2} + \frac{1 - \frac{v'^2}{c^2} + r\frac{w_r'}{c^2}}{r^2\left(1 - \frac{v_r'}{c}\right)^2}\left[\cos rx - \frac{v_x'}{c}\right];$$

$$K_y = \cdots. \qquad \text{(XV)}$$

The distance r is directed along ee' and the quantities v', w' must be taken at a previous time t' such that the wave emanating from e' at that time reaches e at the time t. Since the coordinates $x'y'z'$ of e', xyz of e and their derivatives, the velocities and the accelerations are all definite functions of time, the instant of emission t' is given as an implicit function of t by the equation:

$$c^2(t - t')^2 = [x(t) - x'(t')]^2 + [y(t) - y'(t')]^2 + [z(t) - z'(t')]^2 \cdots. \qquad \text{(XV}a\text{)}$$

[10] Tr. See Ref. A, p. 155, equation XIV.

[11] This expression has been given by K. Schwartzschild, Göttinger Nachr., *Math. Phys. Klasse*, 1903, p. 126. Tr. It is obtained by following the procedure indicated in § 4, p. 137, of Schwartzschild's second paper.

In the case where the velocities are very inferior to that of light, and where the changes are not too rapid (quasi-stationary states), that is, in the majority of cases which present themselves in electrodynamics (with the exception of Hertzian oscillations and the experiments of Kaufmann on β-rays), a function such as $f\left(t-\frac{r}{c}\right)$ can be expanded in a Taylor series:

$$\left(t-\frac{r}{c}\right)=f(t)-\frac{r}{c}f'(t)+\frac{r^2}{2c^2}f''(t)-\cdots$$

and the terms having a factor $\frac{1}{c^2}$ can be neglected. We can thus obtain the elementary action exerted by e' on e in the form of action at a distance, ρ being the *actual* [12] *distance* and v, v', w' the *actual* [12] velocities and accelerations of e and e':

$$F_x=ee'\left\{\frac{\cos\,\rho x}{\rho^2}\left(1+\frac{v'^2-3v_\rho'^2-2v_xv_x'-2v_yv_y'-2v_zv_z'}{2c^2}\right)+\frac{v_x'v_\rho}{\rho^2c^2}-\frac{w_x'+w_\rho'\cos\,\rho x}{2\rho c^2},\qquad F_y=\cdots.\right\}\quad\text{(XVI)}\ [13]$$

This form is particularly suited for the discussion and for a comparison with the classical theories.

What has been said here about the theory of Lorentz is sufficient to enable the reader to follow the criticisms to be made on this theory.

First, as stated above, Lorentz considers only the particular integrals XII and XIII of the partial differential equations IX, X and XI; by hypothesis he removes all other integrals. Let us note the fundamental importance of this restriction. Contrary to what is true of mechanical phenomena, electromagnetic

[12] Tr: e.g. that corresponding to the time t.

[13] Tr. The development of this formula is given in Ref. A, pp. 228–229. It is obtained by an expansion of $x'\left(t-\frac{r}{c}\right)$ and $v_x'\left(t-\frac{r}{c}\right)$, etc., in series like f, followed by a transformation of the potentials as given by Liénard and Wiechert, and finally substitution is made in (VI).

phenomena are in general irreversible on account of the radiation. But the Lorentz equations do not change when we change the sign of the time; they express the condition of reversibility. On the other hand, in the retarded potentials and in the elementary actions XIV, the positive and negative directions of time play quite different roles. There *a velocity is introduced, the sign of which it is, by hypothesis, impossible to change;* namely, the velocity with which the waves constantly travel away from the body which emits them; and this is the origin of the irreversibility of the electromagnetic phenomena. Now, it is easy to verify that the system IX, X and XI admits of an infinite number of other integrals besides XII and XIII, satisfying the conditions of continuity and the conditions at infinity. The general integral allows, in fact, the introduction of two arbitrary functions. Among those integrals some correspond to convergent waves, which, instead of depending on the argument $\left(t - \frac{r}{c}\right)$ depend on the argument $\left(t + \frac{r}{c}\right)$ but have otherwise the same form as XII and XIII. Hence, these waves come from infinity and converge entirely towards the points where the electric charges are located, in the same way as the retarded potentials correspond entirely to waves which travel away from those points. Other integrals, moreover, can be considered as corresponding to waves which converge or diverge from points in the ether. Now, it can be shown that all these waves are physically absurd, since they entail the possibility of perpetual motion. In particular, if in XII and XIII we change $\left(t - \frac{r}{c}\right)$ into $\left(t + \frac{r}{c}\right)$ that is, if we change the sign of c, it can be readily verified that Poynting's radiant vector changes sign. Since the first solution corresponds to bodies which lose energy when they radiate (that is, when they contain particles endowed with non-uniform motion), the second corresponds to *a gain in energy*, and this energy is not lost by any other body, it comes from

infinity and is furnished entirely by the ether, which is thus an inexhaustible reservoir of electromagnetic energy; but a system which under such conditions gathers energy is one of perpetual motion. Hence, *the equations of Maxwell and Lorentz admit of an infinite number of solutions satisfying all the requirements of the theory but incompatible with experience.*[14]

It is therefore quite indispensable to make supplementary hypotheses, whether on the initial state or on conditions at infinity, so as to remove in an absolutely general and complete manner all solutions other than XII and XIII. But this seems to be impossible without giving up the very foundations of the theory. I have shown, in fact (Reference A, p. 166 et seq.), that the only admissible and sufficient condition [15] is that the formulas XII and XIII be verified for the initial state at the time $t = t_0$ and for the consecutive instant $t_0 + dt$. All other hypotheses hitherto proposed, in particular the hypothesis of Poincaré, Abraham and others, which makes the fields zero at a great distance for the initial time t_0, are inadmissible; the last named, for instance, leads to *exclusively convergent* waves for $t < t_0$. But to require that at the instants t_0 and $t_0 + dt$ the field shall correspond with formulas XII and XIII, is to express a condition which has no sense, especially along the ideas of Maxwell. He, in fact, regarded as an essential point in his doctrine that it shall not entail the consideration of the elementary actions and the origin of the field, and that it needs to occupy itself only with the immediate environment of the point under consideration. It is seen that this cannot be right and that, in order to eliminate the physically impossible solutions from those equations, it is necessary to adopt *a priori* the formulas for the retarded potentials, which emphasize the elementary actions as in the classical theories, and to verify that they satisfy the equations. Hence the elementary actions can thus completely replace the differential equations, while the opposite

[14] Tr. This is explained more fully in Ref. A: First Part, § 3.

[15] *Tr.* That is, the only condition for the removal of solutions other than XII and XIII.

proposition is not true. *The differential equations are intrinsically unfit to express the totality of the laws of propagation of the electric and optical actions.*

But let us adopt the formula for the retarded potentials. What is then the significance of the vectors E and H, which seem to be such essential conceptions in the theory? I maintain that these vectors eliminate themselves completely and are useful only as a mathematical tool in some particular cases.[16] In fact, even without knowing the significance of E and H, we can by hypothesis integrate the equations by means of XII and XIII; we need only introduce these values in VII, VIII and VI in order to obtain the quantity F (or the mechanical force exerted on unit charge) expressed as the sum of elementary actions exerted by the other elements of the charge. Now, F eliminates itself, for the state of rest and movement of the electron or of the electrified system is, by hypothesis (whether or not real masses exist), determined by d'Alambert's principle:

$$\Sigma \left(m \frac{d^2x}{dt^2} - F_x - P_x \right) \delta x + \cdots + \cdots = 0, \quad \text{(XVII)}$$

where we suppose F_x replaced by its value given by the elementary actions, and where P represents the resultant of the non-electrical forces. Now, it is only this movement or this state of rest, which form the object of experiment; the field in the pure ether never plays any role. In fact, in order to know the field at any point, we must place an electric charge there. It would be different if E and H, in modifying the ether, set it in motion or changed its properties, as assumed by Maxwell. We could then, for instance, by means of the interference of light waves, exhibit these modifications without placing a charge at the point under consideration. But numerous experiments made for this purpose

[16] Ordinarily E is defined as a mechanical force exerted on unit charge at a point, this charge being at rest relative to the ether, but we cannot bring about this condition and the definition should therefore be rejected. In reality we never observe anything but F and deduce E and H from VI; these vectors are defined by the equations, whatever our point of view.

have always given negative results, and moreover, the hypothesis of such movements has not led to any mechanical explanation of the electrodynamic laws. Lorentz, and with him many physicists, have therefore been led to abandon this idea.

We see then that if we have regard only to the facts, we can do entirely without the notions of electric and magnetic fields, as well as with the partial differential equations and their conditions of continuity; these elements are insufficient, as shown above, to find the solution. The elementary actions, or, to be more exact, equation XVII, and not the partial differential equations, form the complete and adequate expression for Lorentz' theory. Those actions have, moreover, the great advantage that they contain only relations between time and space together with certain invariable coefficients, called electric charges. The notion of force is completely eliminated.

Now, as shown by Schwartzschild (referred to above), the elementary actions link up in the most direct manner with the classical theories. Clausius suggested a formula which, under the hypothesis of action at a distance, expressed the action of one charge in motion on another (analogous to Weber's famous formula, only that it dealt with absolute motion). It is sufficient to introduce in that formula the law of the propagation of light in order to obtain Lorentz's formula. Considering that the same idea had already occurred to Gauss and Riemann, it is surprising to see how much, in this respect, science has deviated from the development along a straight course, and along what tortuous paths and through what difficulties in logics it has had to pass in order finally to end so near its point of departure.

But, in following those tortuous paths, science gained in several ways. She learned to conceive light as an electromagnetic phenomenon and this fundamental concept modified the science of optics itself profoundly. For all that was said above concerning electric phenomena applies to optics. The ether as well as the partial differential equations can be regarded as of secondary importance; in reality, that which we observe

are only elementary actions between ions or electrons in the source of light, in the optical apparatus and in the retina or the photographic plate. All the optical phenomena are derived from those actions in accordance with the principle of superposition.

The ether, which seemed to be an essential concept in the theory, has remained in hiding in all experiments and has been forced to recede from step to step, until its role is reduced to that of a system of absolute coordinates, that is, coordinates which are independent of ordinary matter, and with respect to which the velocities of waves and electrons must be measured. Is it any wonder that experience has rejected this mathematical phantom, and that in spite of Lorentz's formulas, absolute motion does not now seem to play any role in physics?

Let us mention, in passing, other objections to the notion of the ether [17] such as it is generally conceived by modern physicists. The distribution and the movement of energy remain in a large measure arbitrary; there exists even several *simple* solutions of the problem (Ref. A, pp. 172–179). Moreover, in suppressing the idea of motion in the ether, the principle of action and reaction has been therewith suppressed, even although other pictures which can be formed on the propagation of waves may permit to save this principle, as shown below. Finally, the notion of the field is inapplicable to gravitation (Ref. A, p. 179), for as Maxwell himself has remarked, the ether would have to be in unstable equilibrium due to the negative sign of gravitational energy. Hence the field concept cannot constitute a general basis capable of replacing mechanics.[18]

[17] Tr. See also a paper by Ritz in *Scientia*, Vol. III, 1908: "Rôle de l'Ether en Physique."

[18] I cannot here explain the difficulties to which the notion of an elastic ether gives rise, or show how superficial is the analogy between the equations of Maxwell or of optics and those of elastic bodies. There can hardly be any doubt that this analogy raised, in the mind of Maxwell and many others, the vain hope of a mechanical explanation of electric actions. The reader will find a presentation of this question in the memoirs referred to above and in the second edition of Poincaré's *Electricité et Optique* (4th. partie), Paris, 1901.

The formulas XIV and XV for the elementary action of point charges, which, as explained above, sum up the theory of Lorentz, contain only the absolute velocities, either explicitly or through the law of propagation, XV*a*. But since, so far, only relative velocities have entered into and influenced the experiments, it is evident, *a priori*, that it must be possible, without coming into conflict with the observed facts, to make important changes in these formulas in so far as the velocities are concerned; that is, the formulas are in a large measure *hypothetical*. In order to define these changes in a precise manner, consider first the quasi-stationary phenomena to which formula XVII applies. It has not been possible hitherto to observe the electrodynamic or electromagnetic actions depending upon the velocity of ions except in the case where the magnetic field is due to a current, which is closed or practically closed, and where certain velocities are negligible as compared with others. The experiments on induction by movement, on the cathode rays, those of Rowland, Eichenwald, etc., belong in this category. It is found in that case that in equation XVI:

1: The terms in v'^2 and $v_\rho'^2$, introduced by the development in series of $f\left(t - \frac{r}{c}\right)$, remain without sensible influence.

2: For the terms:

$$\frac{ee'}{\rho^2 c^2}\{-[v_x v_x' + v_y v_y' + v_z v_z']\cos \rho x + v_x' v_\rho\} = f_x, \qquad \cdots$$

we can substitute:

$$ee' \frac{\cos \rho x'}{\rho^2 c^2}\{-(v_x v_x' + v_y v_y' + v_z v_z') + 3 v_\rho v_\rho'\} = f_{1x}, \qquad \cdots$$

(which corresponds to Ampère's formula for the action between two elements of current). More generally, we can add to these terms the difference $A(f_x - f_{1x})$, $A(f_y - f_{1y})$, $\cdots$ where A is an *arbitrary* constant, without affecting by this addition the accord between theory and experiments. Finally, we can complete

these expressions by terms in v'^2, v^2, etc., in such a manner that they contain nothing but the *relative velocities* and so that action and reaction are equal, which is not the case with f. Put

$$v_x - v_x' = u_x, \text{ etc.}; \qquad u_x^2 + u_y^2 + u_z^2 = u^2$$

and denote by k an arbitrary constant, then the most general expression for the electrodynamic terms, that contain only the *relative velocities*, becomes: [19]

$$f_{2x} = \frac{\cos \rho x}{4\rho^2 c^2}\left[(3-k)u^2 - 3(1-k)u_\rho^2\right] - \frac{(k+1)u_x u_\rho}{2c^2\rho^2}.$$

But we might also assume that only the rotation of movable electrons in a current can produce a magnetic field. We can see this, if we think of the fact that the action of the magnetic field has not been observed except when it is due to closed currents, there exists in all such cases a magnetic potential equal to the solid angle subtended by the current C (considering only the case of linear currents). Now, the surface of a polygonal figure traced out on a sphere is expressed by the sum of the angles which are formed by each one of the sides of the polygon with the following. For a continuous curvature of the contour, these angles become the angles of contingence of the spherical curve and they can be expressed by means of the radius of curvature of C and its direction relative to the radius vector.

[19] Tr. This expression, then, gives the velocity terms in Ritz' formula for the elementary force referred entirely to the instant t and comprising only relative velocities. In the 2d part of Reference A, pp. 225–227, Ritz shows that this expression represents in the most general way the two laws proposed by Weber and Riemann. It applies even to the action of closed currents on free ions, for instance cathode rays, and to the action which electric charges in motion exert on a magnet. The significance of the constant k is discussed in § 17 of Reference A under "General Remarks," pp. 272–274, and a more complete discussion of the same or rather an analogous constant is found in the papers by Helmholtz (see: *Wissenscht. Abhandl.*, Vol. I, p. 567, where, apparently, this constant was first introduced; see also p. 573 and p. 688). Helmholtz introduced his k in order to simplify his generalized expression for the electrodynamic potential, which besides k contained a velocity $\frac{1}{A}$ which, by measurement, was found to be equal to the velocity of light.

The molecular hypotheses enable us to express this curvature either by the acceleration of the electron, or, for an unsymmetrical electron, by its rotation. We obtain thus an entirely new resolution of the action of a closed current in elementary actions, which can be considered valid for any element of current, closed or not, and which, by integration along the current, will furnish the magnetic potential in all cases. The magnetic field produced by an element of current being thus determined, the force exerted on a moving charge e will be, as in Lorentz' theory: $\frac{e}{c}(v_y H_z - v_z H_y)$, $\cdots$, $\cdots$ where v denotes the velocity of the charge *relative* to the element.

In short, as regards the terms depending on the velocities, we are not better informed than at the time of Weber and Helmholtz.

The case is quite different for the term in XVI which depends on the accelerations:

$$\varphi_x = -\frac{ee'}{2\rho c^2}[w_x' + w_\rho' \cos \rho x], \qquad \varphi_y = \cdots.$$

It can be said that all experimental efforts, all that we know of electric oscillations, of the equations of light, of induction, whether in open or closed circuit, bear solely and entirely on this term. It can be decomposed into two other terms:

1:

$$ee'\left[\frac{-w_x' + w_\rho' \cos \rho x}{c^2 \rho}\right] = \psi_x, \qquad \cdots$$

which, except for quantities in $\frac{1}{c^3}$, is nothing but the term:

$$\frac{-w_x' + w_r' \cos rx}{c^2 r\left(1 - \frac{v_r'}{c}\right)^2}, \qquad \cdots$$

of equation XV, and plays the role of Fresnel's vector in optics. On this term depends all the phenomena exhibited by Hertzian

oscillations at great distances from the source. On the other hand, this term does not play any role in the induction of a closed circuit, for it can be written:

$$-\frac{d}{dx}\frac{w_\rho'}{c^2}, \qquad -\frac{d}{dy}\frac{w_\rho'}{c^2}, \qquad \frac{d}{dz}\frac{w_\rho'}{c^2},$$

so that the expression for the electromotive force:

$$\psi_x dx + \psi_y dy + \psi_z dz,$$

becomes a total differential and vanishes for a closed circuit. The second term is:

2:

$$\frac{ee'}{2c^2\rho}\left[w_x' - 3w_\rho' \cos \rho x\right] = \chi_x,$$

which comes entirely from the development in series, and consequently from the fact that *the velocity of propagation is finite.* It determines the phenomena of induction in a closed circuit and the electric forces in the immediate neighborhood of a Hertzian oscillator (together with the electrostatic term $\frac{ee' \cos \rho x}{\rho^2}$ about the form of which there is no doubt).

But it is important to note that *a law of propagation, according to which the centre of a wave, emitted by an ion at the time* τ, *remains always in straight and uniform motion with a velocity equal to that of the ion at the time* τ, *does likewise by development of the electrostatic term yield the expression* χ_x, *and this result can even be generalized.*[20] We have thus no right to conclude that the centre of the wave is at rest as assumed in the theory of an immovable ether.

The reaction of an electrical system on itself, when there is acceleration, that is, the expression for electromagnetic mass for small velocities, depends exclusively on φ. The existence of such a reaction, therefore, cannot be in doubt; it is absolutely

[20] Tr. This is seen from Reference A, equation (11) of the Second Part, p. 224.

independent of any uncertainty as to absolute motion, or as to the electrodynamic terms and the law of propagation.

Let us now return to the general formulas XIV and XV. *We may in this latter delete*[21] *all the* v' without any experimental fact in optics or electrodynamics being thereby sensibly modified; only $-\frac{v_x'}{cr^2}$ must be retained. This term, of the first order relative to the velocities, is however of no importance in optics or in Hertzian oscillations, and in XVII no term of this kind remains. Another law of propagation, not entailing an absolute system of coordinates, would render it useless.

The linear term of XIV in v (the electromotive force properly so called), contains $\frac{1}{c^2}$ as a factor and is of importance only in the quasi-stationary phenomena; we have seen how uncertain its form is.

The last statements are subject to two restrictions regarding terms of an order higher than the second. The pressure produced by light corresponds to such a term, depending on the product of accelerations and velocities, but its form remains very uncertain. Moreover, the experiments of Kaufmann on β-rays of radium exhibit the totality of these terms in XIV. Unfortunately, no conclusion can be drawn from that, for we may add to the formulas of Weber, or Clausius, and Riemann, or finally to f_{2s} an infinite series of terms depending on powers higher than the second and pairs of velocities, divided by c to the corresponding powers. These terms will be of significance only when the velocities are comparable to c, as in Kaufmann's experiments. Any one of these theories by a suitable choice of the series, could be made to satisfy the observed facts. This shows (Ref. A, pp. 189–197, 260–270) that if there exists an electromagnetic inertia reaction—and this seems beyond a doubt—the variability of this reaction with the velocity is, on

[21] Tr. This word is written "tracer" in the French text, but is believed to be a misprint for "éraciner."

the other hand, hypothetical. It can be deduced from Kaufmann's experiments only if we adopt *a priori* the hypotheses of Lorentz on absolute motion and on the f_s forces. *The theory of variability of electrodynamic mass rests thus on the weakest points in Lorentz' theory. We can explain the observations quite as well, and perhaps better, by appropriate changes in the velocity terms of the expression for the force, introducing merely relative movements.*

It is hardly necessary to add that what little we know of the molecular forces does not permit us to assert that the known laws of electricity are valid at all distances, however small. In reality, we deal always with the law for point charges, and we know of no experiment where the restrictive conditions, which it implies, are not fulfilled.

We have finally to ask ourselves what we have learned experimentally as regards the law of propagation, or, in other words, as regards the equation which connects the instant of emission t' with the instant t, when the action takes place. In Lorentz's theory, the wave emitted by an electron in uniform motion at time t' is at any subsequent time of spherical form, the centre of the sphere remaining at the point in the ether where the wave was emitted, while the electron moves on. This hypothesis, then, introduces absolute motion and if it is admitted that future experiments will not bring this motion into evidence any more than they have done in the past, it should be rejected and the movement of propagation of light should be considered purely relative and dependent on the movement of the source—unless we want to abandon kinematics and the notion of time as done by Lorentz and Einstein.

The principle of *relativity of motion*, in its *classical* form, requires that the waves emitted by a system in uniform motion, and free from any appreciable external influence, shall move with the system, so that the centre of each spherical wave shall continue to coincide with the electron that emitted it, while the radial velocity is universal and equal to c. If the motion of the

electron is non-uniform, the principle of relativity does no longer determine the velocity with which the centre of the wave is displaced; but in any case, this velocity must be constant (otherwise there would be action at a distance between the electron and the emitted wave). It would indeed no longer be possible to conserve the picture of the "ether" or of "waves in an elastic body" for this law of propagation, but, if all the same we want to conserve it, and therewith the partial differential equations, we must adopt the new hypothesis of the Lorentz-Einstein transformation, which, to tell the truth, changes the conditions of the problem so profoundly that the picture of an "ether" or an "elastic body" becomes quite inapplicable. Moreover, according to the views of Lorentz and Einstein, the propagation of light no longer presents a mechanical picture at all. On the other hand, the law of propagation, which we have enunciated above, represents simply a picture of particles emitted in all directions with the same radial velocity and continuing in uniform motion; in this respect, then, it approaches the emission theory of Newton. I have shown (Ref. A, Second Part) that in assuming this law to be true whatever be the motion of the electron, and admitting that these fictitious particles act on electric charges with which they come in contact, it is easy to construct an infinity of electrodynamic theories in perfect accord with the experiments, except insofar as the optics of bodies in motion is concerned. Here, experience, interpreted in terms of the atomic conception of electricity, which we have adopted hitherto, enunciates the following unique and simple result:[22]

When a luminous ray sets the ions of a body in vibration, and this in turn emits new waves, the centres of these waves move, not with the velocity of this body (as according to our hypothesis), but with the velocity of the source of the light.

[22] This theorem is readily verified by following closely the demonstration of Lorentz in *Versuch einer Theorie der elektr. u. opt. Vorgänge in bewegten Körpern*, Leiden, 1895.

Now, this is what the principle of action and reaction would suggest. In reality, this principle is violated by our hypothesis, since to the action of our fictitious particles (if we may use this image) on the ions, there is no corresponding reaction of the ions on the particles. We ought as in the theory of Lorentz, to attribute a certain *momentum* to the radiant energy, which is much more natural when we consider this energy as projected than when we consider it to be propagated. Then the initial velocities of the fictitious particles emitted by an ion will be determined by the principle of *the conservation of momentum*, or principle of reaction. In optics, the whole radiant energy comes from the source, the screens or optical apparatus furnish no additional energy, so that it is natural to assume that the principle of reaction, whatever be its precise expression, will cause the velocity of the fictitious particles, emitted by the ions of the screen, etc., to be entirely determined by the velocity of the source of light.[23]

Evidently, whenever a general and simple theory based on new points of view, fails, there is ground for studying the question in all its aspects, and in particular, not to let one be arrested [24] by the necessity of adopting entirely new kinematics and dynamics, as required by the Lorentz-Einstein hypothesis.[25] But it is important to note that *nothing*, so far, forces us to consider this latter hypothesis as true, or even as particularly probable, and I think it would be deplorable for the science of physics if there could not be found a more simple method of representing

[23] Tr. Hence Ritz finally accepts the theory that light after reflection still retains the vector velocity of the source, combined with c.

[24] Tr. The French word is arrêter. Presumably Ritz means that we should not be afraid to adopt, if necessary, entirely new kinematics, etc.; but he goes on to say that *in the present case* nothing forces us to adopt the Lorentz-Einstein hypothesis.

[25] Besides, it must be noted that the Lorentz-Einstein theory is likewise, in part, only at the stage of a program; the principle of d'Alambert, and more generally the classical dynamics of systems, are incompatible with that theory, as Einstein himself has remarked, and nothing has appeared as yet to take the place of these fundamental principles. In the author's hypothesis, on the contrary, they continue to exist without change.

the electric and electrodynamic laws than that which consists, first, in assuming absolute coordinates, then in writing down a system of 11 equations, of which 9 are expressed in partial derivatives; in integrating those equations by means of hypotheses, which must be rejected afterwards in order not to introduce impossible solutions or in order to exclude other possible solutions, and finally complicating the formula so obtained, already very long, by a transformation which destroys the principles of kinematics and the only object of which is to get rid of the absolute system of coordinates, so unhappily introduced at the beginning. Yet, these are not the only reasons why this theory does not seem to me satisfactory. It is asserted that the equality of the ratio between the units of electricity and the velocity of light, c, finds its explanation in the theories of Maxwell and Lorentz. Now, the complexity of the former theory renders it difficult to obtain a clear view of the manner in which this result is attained. But let us consider the equations of Lorentz. Here the velocity c figures in various manners, and it is not difficult to see that in retaining c in the partial differential equations, but changing $\frac{1}{c}$ to $\frac{1}{c'}$, in the expression VI for the force, where c' is different from c, we neither modify the velocity of propagation, nor the units or the energy.[36] But we do change the ratio between the electrostatic unit and the electrodynamic unit, and the theory presented in this form contains nothing that could make us understand why the coefficient of $(v_y H_z - v_z H_y)$ should be precisely equal to $\frac{1}{c}$. It is so chosen, because it is demanded by experiments, exactly as in the formula of Weber, Clausius, etc. It is only when it is required that Hamilton's

[36] The term in $\frac{1}{c^2}$, in fact, always gives zero amount of work. It must also be remembered that H, as stated above, is defined by the theory itself; if it were defined *a priori* as a force (expressed in gauss units) acting on the pole of a permanent magnet, the coefficient of ρv would in turn become an empirical coefficient and the conclusions would be the same.

principle shall be applied, in a certain special form, that the coefficient $\frac{1}{c}$ is found *a priori*. But, this principle, as it is used by Lorentz, is distinctly different from the principle in the ordinary sense. The variations are taken in quite another way than in the case of fluids, for instance, and moreover, as shown by Schwartzschild, the principle can be applied in different ways. One of these determines directly the elementary forces, without considering the field, and we prefer that from the point of view presented in this study. Lagrange's function has the same form (except for the propagation) as that from which Clausius started:

$$\int\int \frac{dE dE'}{r}\left[1 - \frac{v_x v_x' + v_y v_y' + v_z v_z'}{c^2}\right],$$

where dE and dE' are elements of charge and where v and v' must be taken at appropriate instants.

If we change $\frac{1}{c^2}$ into $\frac{1}{c'^2}$ the formula no longer satisfies the experimental results, the ratio between the units will no longer be c, but the principle of least action continues to apply and the velocity of propagation will remain equal to c. In that case it is the partial differential equations that are not satisfied.

Summing up, it is seen that this remarkable relation does not result from Lorentz's theory except in a manner so indirect that it is reduced to a choice of coefficients, as with Weber and Clausius, unless the condition due to the principle of least action is taken into account. In didactic works, as for instance, in the treatise by Abraham and even in the presentation given by Lorentz of his theory in the memoir referred to above, this condition is not mentioned, and seems to be regarded as absolutely secondary.

Gauss, in his famous letter to Weber, indicated that undoubtedly the electrodynamic terms resulted exclusively from the finite velocity of propagation, conceived in accordance with a properly selected law, by means of an expansion in series

such as considered above, which, as shown, introduces effectively the velocities and the accelerations together with coefficients that depend upon c The relation between the ratio of the electromagnetic units and the velocity of light would then have an immediate significance. With Maxwell, the electrodynamic terms, depending on the vector potential, enter in quite another manner; the profound conception of Gauss has not been realized by Maxwell and Lorentz (it is important to assert this, for the contrary has been often maintained). To that, perhaps, will belong the future of electrodynamics.

III. Ritz's New Theory

Ritz fully realized the enormous difficulties in developing a new theory of electrodynamics and he emphasized repeatedly that the theory as presented in the Second Part of his great Memoir [27] was entirely preliminary and incomplete, that it was purposely given such a form as to be specially adapted to a comparison with Lorentz's theory, and that it would require considerable modification.

The following is essentially an abstract of that preliminary outline, in which it is shown how the formula for the elementary actions can be modified without ceasing to be in accord with the experimental results.

Ritz discarded the hypothesis of the ether and therewith the system of absolute coordinates, and the idea of absolute motion. He did not, however, revert to the theory of action at a distance. He adopted unaltered Lorentz's theory of electrons and the principle of superposition, but did not make use of such notions as polarization, the electric vector, and the magnetic force. He used only time, space and electric charges, the latter playing the same part as masses in Mechanics. In a certain sense this was a mechanical theory of electricity.

Ritz gave to the law of propagation of electrodynamic actions a very simple kinematical form, borrowed from the emission

[27] *Annales de Chimie et de Physique*, 1908.

theory of light, which is in strict accordance with the classical law of relativity. Fictitious luminous particles, supposed to be constantly expelled from a source in all directions, continue to move with uniform velocity in straight lines. They fill constantly a sphere, the centre of which has the same movement of translation as that of the source at the instant of emission. If this velocity is constant, the centre of the sphere will continue to coincide with the source. The velocity of the particles combines vectorially with the velocity of light, and remains invariable, whatever be the later movement of the source and *even when the particles pass through ponderable bodies or electric charges*. The waves in the ether are replaced by a distribution of the emanation, which is periodic in space and time.[28]

This hypothesis was applied to electrodynamic action, but, as stated above, it is quite provisional. It is contrary to the principle of action and reaction and is not applicable to the optics of moving bodies, but was preferred by Ritz in order to keep as close as possible to Lorentz's hypothesis. In this way he was able to show point for point that it was possible by dealing only with the elemental force to obtain the same results as Lorentz, but otherwise he was prepared to adopt an emission theory, which conformed best with experience. He admitted in fact[29] that it would be more natural in the emission theory to assume that radiant energy possesses momentum, whereby it becomes possible to comply with the principle of action and reaction. He even discerned in this principle the possibility of obtaining the electrodynamic terms, which, in the expression for the elemental force, depend upon the velocity and the acceleration.

In principle Ritz adopted the theory developed by Schwartzschild,[30] according to which the elemental force F, exerted by an electric point-charge e' on another charge e, has one component par-

[28] Ritz: *Scientia*, Vol. III, "Rôle de l'Êther en Physique," 1908.

[29] Reference A, p. 151. See also p. 245 of the present paper.

[30] Gött. Nachrichten, *Math. Phys. Kl.*, 1903: "Zur Elektrodynamik," II, p. 135.

allel with the line joining the two charges, another parallel with the velocity v' of e' and a third parallel with the acceleration w' of e'. He modified the expression for F by substituting the relative velocity u for the absolute velocities v and v'. In order to take into account the entire domain of electrodynamics, he considered first the case where the velocities and accelerations are relatively small as in classical electrodynamics and for cathode rays, etc. Second, he considered the case where the accelerations are great, but the velocities very small relative to c, as in case of Hertzian oscillations, and finally the case where the velocities are comparable to c, but the accelerations small, as in Kaufmann's experiments on the β-rays of radium. Ritz showed that the forces and movements and consequently the work, obtained by his theory, were quite the same as those found by Maxwell's theory for all the phenomena observed up to that time.

In the first case, where the velocities and accelerations are small, referred to by Ritz as the phenomena of "slow variation," all functions of $\left(t - \frac{r}{c}\right)$ are expanded in Taylor's series and an equation for the elemental force is obtained in which all quantities are referred to the time t and the charge e. The quantity r is here the radius of the spherical wave emitted from e' and reaching e at the time t. Denoting now by ρ the *actual* distance between the charges at the time t and introducing an arbitrary constant k, the following general formula for the elemental force is obtained:[31]

$$F_x = ee'\left\{\frac{\cos\rho x}{\rho^2} + \frac{\cos\rho x}{4\rho^2c^2}\left[(3-k)u^2 - 3(1-k)u_\rho^2\right] - \frac{(k+1)u_x u_\rho}{2c^2\rho^2} - \frac{w_x' + w_\rho'\cos\rho x}{2c^2\rho}\right\}, \qquad F_y = \cdots.$$

This is the formula of which certain parts were discussed separately in the translated paper, Chapter II. It will be noticed that the expression on the right-hand side consists of three terms or groups of terms; first, an electrostatic term, defining an action

[31] Reference A, p. 225, equation (13).

along ρ; second, a group of terms depending only on the relative velocity of the charges, and third, a group of terms depending only on the accelerations of e'. Here not only ρ but also u and the density are referred to the time t and the expression, therefore, has the convenient form corresponding to action at a distance.

When the accelerations are very great, as in Hertzian oscillations, the expansion in series of functions of $\left(t - \frac{r}{c}\right)$ cannot be employed. The formula for the elemental force takes a different form, and the terms in u can be neglected, leaving only the electrostatic term and the term in the accelerations. It is shown that the resulting force in that case is the same as that obtained according to Lorentz's theory.

When the velocities are very great as in Kaufmann's experiments, Ritz's analysis leads to the conclusion that instead of assuming absolute motion with variability of mass, the results of those experiments may be interpreted equally well by considering the movements as relative and the masses as constant. It is assumed that for great velocities the electrodynamic forces are not simple linear functions of the velocities, as according to Lorentz's theory, but that they take on a more complicated form. In that case it cannot be asserted from these experiments that the mass of the electrons is of electromagnetic origin, although this remains possible.

Ritz's theory accounts easily for the aberration of light and for the results of the Michelson-Morley experiments with light from terrestrial sources. It was finally applied to gravitation,[32] which Ritz suggested might be due to a rotation of the positive charges in the atoms, arguing that the magnetic phenomena required that the charges must be in motion. When two bodies, each consisting of a great number of atoms, act on one another, the velocities and accelerations produce electrodynamic forces, which, although the axes of rotation have all possible directions,

[32] Reference A, p. 267, and *Scientia*, Vol. V, No. X–2, April 1909: "Die Gravitation," pp. 241–255.—Ritz, *Gesammelte Werke*: "Das Prinzip der Relativität in der Optik," p. 517.

do not on the average vanish. Provided that relative and not absolute movements are used and provided we adhere to classical kinematics, the expression for the elemental force will contain terms, the mean value of which is different from zero. These terms are of an order higher than the second and are associated with unknown coefficients. They yield a resultant force inversely proportional to the square of the distance between the two bodies, and directly proportional to the number of rotating charges which they contain, that is, to the masses. Since the terms involved are of high order, perhaps 4th. to 6th., it is clear that the gravitational action of two atoms on one another must be small, as it is known to be. Summing up, Ritz predicted that it would be possible to reduce gravitation to electrical forces, and to determine the gravitational constant from electromagnetic measurements; further, that it would be possible to explain the slow rotation of Mercury's orbit by applying the laws of electrodynamics, when once these laws become known with sufficient accuracy. He developed formulas to that effect.[33] Whatever be the assumptions in detail, gravitation would ultimately depend on the dynamic constitution of the atom.

IV. Concluding Remarks

In view of the great importance of Ritz's work, it might seem strange that no one took up his line of investigation after his death. One reason is probably the fact that certain astronomical observations and laboratory experiments seemed to disprove the emission theory,[34] and another that the great success of the Einstein theory drew the attention and efforts of physicists and mathematicians in a different direction. So long as the nature of light was believed to be one of pure wave motion, there seemed no escape from the Einstein fundamental postulate as regards the invariance of the velocity of light; but now, after the dual nature of light has been established and a definite relation has been shown to exist between the projected quantums

[33] Reference A, pp. 267–271.

[34] R. C. Tolman, *The Theory of Relativity of Motion*, 1917, p. 24.

of electricity and the frequency of the wave motion, it would seem worth while to take up Ritz's theory for renewed consideration. It appears that the time has come when one of the many brilliant minds, now working in the field of mathematical physics, might continue the work so well commenced by Ritz, or at least investigate the possibilities of his line of approach.

In spite of the remarkable progress made in physical science in recent years, several fundamental questions are left in a highly unsatisfactory state.

The ether hypothesis, although abandoned in some arguments is not abandoned in others. It forms the foundation of Maxwell's theory, but is incompatible with Einstein's Theory of Relativity. As expressed by Sir James Jeans:[35] . . . "it seems best to regard it (the existence of an ether or medium) merely as a working hypothesis, to be discarded if it is found to lead to contradictory or impossible results, and to be retained if it proves to be useful as well as self-consistent." It will be admitted that such a theory is unsatisfactory. Although the ether theory has proved useful, it is certainly not self-consistent.

On the other hand, if the medium is discarded, we are left with empty space through which energy and momentum in some way must be transmitted, as evidenced by the existence of electrical, magnetic and gravitational forces. The explanation that such transmission should be possible by pure wave motion through empty space is again unsatisfactory. We are driven by the necessity of logics to assume an emission theory for electromagnetic actions, associated in some way, which we do not yet understand, with wave motion or at least with some wave-like periodicity.[36] From this concept there is only a short step to a theory similar to that proposed by Ritz.

The Restricted Theory of Relativity puts a particular interpretation on certain physical experiments and builds up on this basis a mathematical structure. In this way all difficulties

[35] *Electricity and Magnetism*, 1927, p. 618.

[36] *Gesammelte Werke*, 1911: "Princip der Relativität in der Optik," pp. 509–518.

associated with the ether and with the transmission of radiant energy are indeed eliminated and a remarkable agreement between the theory and physical phenomena is obtained, but it is necessary to accept a postulate which is self-contradictory according to classical mechanics, and it seems legitimate to question whether those experiments cannot be interpreted in a different and less revolutionary way. Ritz's theory promises a solution of this problem.

SYSTEMS OF THREE LINEAR PARTIAL DIFFERENTIAL EQUATIONS OF THE THIRD ORDER [1]

By G. P. Wadsworth

In this paper, we consider the solution of a system of three linear, homogeneous, partial differential equations of the third order.

Previously to this,[2] E. P. Lane, by geometrical consideration, found canonical forms to which the above equations could always be reduced. He also considered in another paper [3] special cases where there were only two differential equations instead of three.

The method used here, however, is an extension of that used by P. Franklin and C. L. E. Moore [4] in considering the solution of two partial differential equations of the second order.

We find that there are two canonical forms to which the original equations can be reduced and we show how the three forms of Lane reduce to these two forms by a particular transformation of variables.

Thus the solution of a system of third order linear partial differential equation reduces to the solution of one of the two canonical forms of which it is a consequence.

In the case of one of the canonical forms, the solutions are linear combinations with arbitrary constant coefficients of eight or fewer definite functions. In the case of the other canonical form, the solutions are certain linear combinations of an arbitrary

[1] The first part of this paper was presented as a thesis for the S.M. Degree at the Mass. Inst. of Technology in 1931, written under the direction of Prof. P. Franklin.

[2] *The Tohoku Math. Journal*, vol. 33, Nos. 1, 2, 1930, p. 12.

[3] *Trans. A. M. S.*, vol. 32, 1930, p. 782.

[4] *Journal of Math. and Physics*, Vol. IX, No. 1, 1930, p. 22.

constant, an arbitrary function and its derivatives where the coefficients of these functions are definite functions which depend upon the original system of equations.

1. Statement of the Problem and Preliminary Reductions

We shall consider three third order linear partial differential equations with two independent variables. Furthermore, we shall assume, first that none of them are consequences of another equation in the same system, and secondly, that there are no second order equations which are consequences of the system. This second assumption also implies that there are no first order equations present for then the second order could be obtained by simple differentiation.

Explicitly, these equations take on the form:

$$
\begin{aligned}
&\sum_{ijk} A_{ijk} Z_{ijk} + \sum_{ij} B_{ij} Z_{ij} + \sum_{i} C_i Z_i + DZ = 0, \\
(1)\qquad &\sum_{ijk} A_{ijk}' Z_{ijk} + \sum_{ij} B_{ij}' Z_{ij} + \sum_{i} C_i' Z_i + D'Z = 0, \\
&\sum_{ijk} A_{ijk}'' Z_{ijk} + \sum_{ij} B_{ij}'' Z_{ij} + \sum C_i'' Z_i + D''Z = 0,
\end{aligned}
$$

where the coefficients may be assumed symmetrical in all pairs of indices and are functions of the independent variables u and v. The subscripts on the z's denote differentiation with respect to these variables with the summation on i and j running from 1 to 2.

Suppose this system of equations has some solution, $Z = k(u, v)$, different from zero. Then the transformation $Z = \bar{Z}k$ may be effected and will of necessity have to be satisfied by $\bar{Z} = 1$. If the equations are homogeneous, they will then lack the term in $\bar{Z}$. We can therefore assume, without loss of generality, that $D = D' = D'' = 0$.

2. Parabolic Equations

We first try to obtain parabolic third order equations as consequences of a linear combination of the original system. So we multiply by three undetermined coefficients, and try to

make the resultant combination parabolic. This will be the case if

$$
\begin{aligned}
&(\lambda A_{112} + \mu A_{112}' + \nu A_{112}'')^2 \\
&\quad = 3(\lambda A_{111} + \mu A_{111}' + \nu A_{111}'')(\lambda A_{122} + \mu A_{122}' + \nu A_{122}'') \\
&(\lambda A_{122} + \mu A_{122}' + \nu A_{122}'')^2 \\
&\quad = 3(\lambda A_{112} + \mu A_{112}' + \nu A_{112}'')(\lambda A_{222} + \mu A_{222}' + \nu A_{222}'')
\end{aligned} \tag{2}
$$

can be satisfied.

This can be seen directly from the fact that if

$$aZ^3 + bZ^2 + cZ + d$$

is to be of the form,

$$a(Z + \alpha)^3,$$

then we may obtain by equating coefficients the two equations:

$$
\begin{aligned}
c^2 &= 3bd \\
b^2 &= 3ac,
\end{aligned} \tag{3}
$$

which are homogeneous and independent.

These equations are necessary conditions for the parabolic equation, but they also can be shown to be a sufficient condition, for we can obtain the original polynomial as a perfect cube by reversing the process, as:

$$\frac{3a}{b} = \frac{b}{c} = \frac{c}{3d} = \frac{1}{\alpha}$$

$$Z^3 + \frac{b}{a}Z^2 + \frac{c}{a}Z + \frac{d}{a}$$

$$Z^3 + 3\alpha Z^2 + 3\alpha^2 Z + \alpha^3.$$

Using Sylvester's Dialytic Method of Elimination, it is possible to remove λ from equations (2) and solve the resulting determinant for the ratio of u/v. Solving back for λ, we have the ratio of $\lambda/u/v$, which will make, the linear combination parabolic and will be satisfied by all solutions of (1).

3. Linear Dependence

If we have two distinct characteristics of these parabolic equations, they can be taken as the new coordinate lines $U = \text{const.}$, $V = \text{const.}$, and thus we are able to obtain two parabolic equations of the form:

$$\begin{aligned} Z_{111} + K' &= 0 \\ Z_{222} + K'' &= 0, \end{aligned} \tag{4}$$

where K' and K'' are functions of the first and second derivatives only.

If we now test for linear dependence by using the generalized Wronskian defined by P. Franklin and C. L. E. Moore:[5]

$$W^2_{S,\infty} = \begin{Vmatrix} Z^1 & Z^2 & \cdots & Z^S \\ Z_1^1 & Z_1^2 & \cdots & Z_1^S \\ Z_2^1 & Z_2^2 & \cdots & Z_2^S \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \end{Vmatrix},$$

where we have S functions of two variables u and v, and the subscripts 1 and 2 denote differentiation with respect to these variables. Reading down each column of the matrix, we have one of the functions, then the two first partial derivatives, the three second partial derivatives, etc., down to the derivatives of the nth order.

We see from equations (4) and the remaining equation left in the system that in the Wronskian W^2_{S4}, six of these rows will be linearly dependent on the remaining eight. And, therefore, the rank of the matrix cannot be greater than eight. It can also be seen from these same equations that all derivatives of higher order than the fourth, are linearly dependent on those of the first, second, third and fourth derivatives. Thus the rank of any Wronskian greater than W^2_{S4} will have the same rank as W^2_{S4}.

[5] L.c.,[4] p. 23.

Therefore, our solutions will be linearly dependent on eight or fewer solutions by the following theorem and its corollary:

"If, throughout some u, v region R the partial derivatives of the S functions Z up to and including those of the nth order exist and are continuous, and the two matrices $W^2_{S,n}$ and $W^2_{S,n-1}$ are of the same rank k, there being in particular some determinant of the kth order in $W^2_{S,n-1}$ which never vanishes in R, then the S functions Z are linearly dependent on precisely k of them."

Therefore, when we have two parabolic solutions in the pencil, we can reduce equations (1) to the canonical form:

$$\begin{aligned} Z_{111} + K' &= 0 \\ Z_{222} + K'' &= 0 \\ Z_{112} + A_{122}Z_{122} + K''' &= 0, \end{aligned} \tag{5}$$

where the K's again are functions of the first and second derivatives only.

4. Existence of One Parabolic Equation

We shall now consider when it is possible to obtain a parabolic solution. There must be either a term in Z_{111} or Z_{222} in one of the equations if there is no second order equations present. If Z_{222}, interchange the coordinates so we have a term in Z_{111}. Solve this equation for Z_{111} and eliminate it from the remaining two.

Therefore, we have:

$$\begin{aligned} Z_{111} + AZ_{112} + BZ_{122} + CZ_{222} + K' &= 0 \\ A_1Z_{112} + B_1Z_{122} + C_1Z_{222} + K'' &= 0 \\ A_2Z_{112} + B_2Z_{122} + C_2Z_{222} + K''' &= 0. \end{aligned} \tag{6}$$

4. (*a*) *Determinant Zero*

We shall first prove that if the following determinant is equal to zero, there is a parabolic equation directly in the pencil under the assumptions that there are no second order equations or any dependence among the equations themselves. We will now

deduce from these assumptions the fact that this determinant:

$$\Delta = \begin{vmatrix} A & B & C \\ A_1 & B_1 & C_1 \\ A_2 & B_2 & C_2 \end{vmatrix} \tag{7}$$

is not equal to zero.

If it is equal to zero, it is equivalent to saying that we may find values of λ, μ, ν, not all zero, satisfying the three equations:

$$\lambda A + \mu A_1 + \nu A_2 = 0$$
$$\lambda B + \mu B_1 + \nu B_2 = 0$$
$$\lambda C + \mu C_1 + \nu C_2 = 0.$$

We cannot have $\lambda = 0$ for then μ times the 2d equation plus ν times the 3d gives a resulting equation with no 3d order terms. If this sum is exactly zero, we have dependence among the equations. If it is not zero, we have 2d order (or 1st) terms present. Hence $\lambda \neq 0$ and λ times the 1st equation plus μ times the second plus ν times the third yields a parabolic equation.

4. (*b*) *Determinant Not Zero*

Now assume that the above determinant is not equal to zero. Take linear combinations of (6) and set up the conditions as in equation (2) for a parabolic solution.

The ratio giving the parabolic equation must have at least one multiplier not equal to zero. Make this multiplier by interchanging equations equal to λ and then make a change of scale so that $\lambda = 1$.

We then have, since $\lambda = 1$ in equations (2):

$$(\mu A_1 + \nu A_2)^2 = T' \tag{8}$$
$$(\mu B_1 + \nu B_2)^2 - 3(\mu A_1 + \nu A_2)(\mu C_1 + \nu C_2) = T'',$$

where T' and T'' are combinations of linear and constant terms.

We also must observe at this point that if $A_1 = A_2 = 0$ (or $B_1 = B_2 = 0$), B_2 times and the 2d equation minus B_1 times the 3d equation, yields a 2d order equation, or if the above is identically zero, we obtain a combination of these same equations.

Because of the fact that determinant (7) is not zero, we shall show that this pair of conics can only intersect at the most in two points at infinity, leaving at least two of the roots finite.

If the ratio of the quadratic terms are equal, that is:

$$(9)\qquad \frac{A_1^2}{B_1^2 - 3A_1C_1} = \frac{2A_1A_2}{2B_1B_2 - 3A_1C_2 - 3A_2C_1} = \frac{A_2^2}{B_2^2 - 3A_2C_2}.$$

We see that neither A_1 nor A_2 can equal zero, without both vanishing. This case has been excluded.

By algebraic manipulation we easily arrive from (9) at the equations:

$$2B_1A_1(A_2B_1 - B_2A_1) = 3A_1^2(A_2C_1 - C_2A_1)$$
$$2B_2A_2(A_2B_1 - B_2A_1) = 3A_2^2(A_2C_1 - C_2A_1).$$

A linear combination of these gives:

$$A_1A_2(A_2B_1 - B_2A_1)(A_2C_1 - C_2A_1) = 0.$$

If $(A_2C_1 - C_2A_1) = 0$, we find that either $(A_2B_1 - B_2A_1) = 0$, or $B_1 = B_2 = 0$ which latter case has been ruled out.

If $(A_2B_1 - B_2A_1) = 0$, we see directly that $(A_2C_1 - C_2A_1) = 0$ since A_1 and A_2 are not zero.

Therefore a necessary and sufficient condition for (7) to hold is:

$$(10)\qquad \frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2}.$$

But case (10) is ruled out by the determinant (7) not vanishing.

Therefore, since equations (10) are never satisfied, the quadratic equations (8) always have a finite solution and we may deduce at least one parabolic equation from our original system.

5. Existence of Two Parabolic Equations

Since we have a parabolic equation, we may write the set of equations (1) in the form:

$$(11)\qquad \begin{aligned} Z_{111} + T' &= 0 \\ A_1Z_{112} + B_1Z_{122} + C_1Z_{222} + T'' &= 0 \\ A_2Z_{112} + B_2Z_{122} + C_2Z_{222} + T''' &= 0. \end{aligned}$$

We are still able to obtain another parabolic except in very special cases, for we have to satisfy equations (2) again and thus we have:

$$(12)\qquad \begin{aligned}(\lambda A_1 + \mu A_2)^2 &= 3(\lambda B_1 + \mu B_2)\\ (\lambda B_1 + \mu B_2)^2 &= 3(\lambda A_1 + \mu A_2)(\lambda C_1 + \mu C_2).\end{aligned}$$

The case where

$$\frac{A_1}{A_2} = \frac{B_1}{B_2} = \frac{C_1}{C_2},$$

has been considered before.

From the second of the quadratic equations of (12) which is homogeneous we are able to obtain the following ratio which we are able to close so that it does not become infinite.

$$\frac{\lambda}{\mu} = \frac{\lambda_1}{\mu_1}.$$

Substituting this ratio in the first of the quadratic equations we arrive at the equation:

$$\mu^2(\lambda_1 A_1 + \mu_1 A_2)^2 = 3\mu\mu_1(A_1 B_1 + \mu_1 B_2)$$

or

$$\mu = \mu_1 \frac{3(\lambda_1 B_1 + \mu_1 B_2)}{(\lambda_1 A_1 + \mu_1 A_2)^2}.$$

We see from this equation that there are two cases where we are unable to obtain a solution and thus unable to obtain another parabolic.

Case I

$$\lambda_1 B_1 + \mu_1 B_2 = 0; \qquad \lambda_1 A_1 + \mu_1 A_2 = 0.$$

In this case:

$$\frac{A_1}{A_2} = \frac{B_1}{B_2}$$

and we have another parabolic equation directly.

Case II

$$\lambda_1 B_1 + \mu_1 B_2 = 0; \qquad \lambda_1 C_1 + \mu_1 C_2 = 0.$$

In this case:

$$\frac{B_1}{B_2} = \frac{C_1}{C_2}$$

and we obtain a new canonical form:

$$\begin{aligned} Z_{111} + T' &= 0 \\ Z_{112} + T'' &= 0 \\ Z_{122} + T''' &= 0. \end{aligned} \tag{13}$$

6. Canonical Forms

We have found in equations (13) and (5), the two canonical forms to which all systems of partial differential equations of the kind considered may be reduced according to whether the system contains one or two parabolic equations.

$$\begin{aligned} Z_{111} + K' &= 0 \\ Z_{112} + K'' &= 0 \\ Z_{122} + K''' &= 0. \end{aligned} \tag{13}$$

$$\begin{aligned} Z_{111} + K' &= 0 \\ Z_{222} + K'' &= 0 \\ Z_{112} + AZ_{122} + K''' &= 0. \end{aligned} \tag{5}$$

We shall now consider in detail how the three following canonical forms, found by Lane, may be reduced to these two fundamental forms. These three canonical forms were found from geometrical considerations.

$$\begin{aligned} Z_{111} + T' &= 0 \\ Z_{112} + T'' &= 0 \\ Z_{122} + T''' &= 0. \end{aligned} \tag{14}$$

$$\begin{aligned} Z_{111} + DZ_{222} + T' &= 0 \\ 3Z_{112} + T'' &= 0 \\ 3Z_{122} + T''' &= 0. \end{aligned} \tag{15}$$

$$\begin{aligned} Z_{111} + T' &= 0 \\ Z_{222} + T'' &= 0 \\ Z_{112} + T''' &= 0. \end{aligned} \tag{16}$$

The subscripts 1 and 2 again denote differentiation with respect to u and v respectively and the T's are functions of the second and first derivatives only.

Equations (14) are identical with the canonical form (13) obtained where there is no linear dependence upon a definite number of functions.

There is linear dependence, however, in the case of equations (15) and (16). In fact equations (16) contain two parabolic solutions directly and from equations (15) also two parabolic solutions may be obtained by multiplying the second equation of the set by $\sqrt[3]{D}$ and the third by the $(\sqrt[3]{D})^2$ and then adding all three. As there are always two distinct cube roots, we may always obtain two parabolic solutions.

Equations (16) are in fact a special case of the first canonical form (5) where A is placed equal to zero. We may also arrive at the canonical form (5) from equation (15) by a change of the independent variables.

Let

$$(17) \qquad \begin{aligned} Z_u &= \lambda(Z_x + w_1 Z_y) \\ Z_v &= \mu(Z_x + w_2 Z_y), \end{aligned}$$

where w_1 and w_2 are two distinct cube roots of D and λ and μ are integrating factors.

Z_u and Z_v denote differentiation with respect to the independent variables of equations (5) and Z_x and Z_y denote differentiation with respect to the independent variables of the set (15).

We shall now show that this relation between the partial derivatives defines u and v as functions of x and y.

As

$$\begin{aligned} \frac{\partial Z}{\partial u} &= \frac{\partial Z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial Z}{\partial y}\frac{\partial y}{\partial u} \\ \frac{\partial Z}{\partial v} &= \frac{\partial Z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial Z}{\partial y}\frac{\partial y}{\partial v}, \end{aligned}$$

we obtain by equating coefficients that the following relations must hold:

$$\frac{\partial x}{\partial u} = \lambda(x, y) \quad \frac{\partial y}{\partial u} = \lambda w_1$$

$$\frac{\partial x}{\partial v} = \mu(x, y) \quad \frac{\partial y}{\partial v} = \mu w_2.$$

Also since

$$\frac{du}{du} = \frac{dv}{dv} = 1, \qquad \frac{du}{dv} = \frac{dv}{du} = 0,$$

we have the four equations:

$$\frac{\partial u}{\partial x}\lambda + \frac{\partial u}{\partial y}\lambda w_1 = 1$$

$$\frac{\partial v}{\partial x}\lambda + \frac{\partial v}{\partial y}\lambda w_1 = 0$$

$$\frac{\partial u}{\partial x}\mu + \frac{\partial u}{\partial y}\mu w_2 = 0$$

$$\frac{\partial v}{\partial x}\mu + \frac{\partial v}{\partial y}\mu w_2 = 1.$$

From equations 2 and 3 of this set we may obtain two total differential equations:

$$\frac{dx}{1} = \frac{dy}{w_1}, \qquad \frac{dx}{1} = \frac{dy}{w_2},$$

whose general solution give u and v as functions of x and y. The integrating factors λ and μ can then be obtained from the two remaining equations of the set.

We may therefore solve (17) for Z_x and Z_y and after carrying out the indicated differentiation we insert this change of variables in (15). Then multiplying these equations in order by λ^3, $\lambda^3 w_1$, $\lambda^3 w_1^2$ respectively and adding we obtain one parabolic solution. Now multiplying the equations by μ^3, $\mu^3 w_2$, $\mu^3 w_2^2$ we obtain the second parabolic equation.

There is one equation remaining which may be obtained by eliminating the Z_{111} and Z_{222} terms from any one of the original three equations by means of the two parabolics obtained.

This third equation is of the form,

$$Z_{112} + AZ_{122} + T = 0,$$

where A can be definitely determined as a function of D and the integrating factors.

Thus we have reduced (15) to the form (5).

Therefore we have the canonical form (5) when we have linear dependence and the canonical form of (13) when we have an arbitrary function in the solution. This completes the transition from the three canonical forms to two.

7. Form of Solution

We shall now consider the character of the general solutions of equations (13), that is, the case where there is no linear dependence. Explicitly these equations are:

$$\begin{aligned} Z_{111} &= aZ_{11} + bZ_{12} + cZ_{22} + dZ_1 + fZ_2 \\ Z_{112} &= a'Z_{11} + b'Z_{12} + c'Z_{22} + d'Z_1 + f'Z_2 \\ Z_{122} &= a''Z_{11} + b''Z_{12} + c''Z_{22} + d''Z_1 + f''Z_2. \end{aligned} \tag{18}$$

From the integrability conditions:

$$(Z_{111})_2 = (Z_{112})_1, \qquad (Z_{112})_2 = (Z_{122})_1,$$

we see that unless c and c' are equal to zero we may obtain a second parabolic solution through the integrability condition for a Z_{222} term would be introduced.

$$\therefore \quad c = c' = 0.$$

7a. General Case

Let us first consider the case where c'' and either f or f' are different from zero.

We now differentiate the first two equations of (18) with respect to v and leave the third unchanged.

Thus we have:

$$
\begin{aligned}
Z_{111v} &= aZ_{112} + bZ_{122} + gZ_{11} + hZ_{12} + kZ_{22} + mZ_1 + nZ_2 \\
Z_{112v} &= a'Z_{112} + b'Z_{122} + g'Z_{11} + h'Z_{12} + k'Z_{22} + m'Z_1 \\
&\qquad + n'Z_2 \\
Z_{122} &= a''Z_{11} + b''Z_{12} + c''Z_{22} + d''Z_1 + f''Z_2.
\end{aligned} \tag{19}
$$

By solving for Z_{22} in the third equation, this term may be eliminated from the first two. We may also eliminate the Z_2 term from the first two of (19) by means of one of the equations of (18). This same equation may be used to eliminate the Z_2 term from the other of the first two equations of (18).

Thus using the resulting two equations of (19) and one of the first two equations of (18) with Z_2 eliminated we have:

$$
\begin{aligned}
Z_{111v} &= pZ_{112} + qZ_{122} + vZ_{11} + sZ_{12} + tZ_1 \\
Z_{112v} &= p'Z_{112} + q'Z_{122} + v'Z_{11} + s'Z_{12} + t'Z_1 \\
0 &= Z_{111} + p''Z_{112} + v''Z_{11} + s''Z_{12} + t''Z_1.
\end{aligned} \tag{20}
$$

Now take $Z_1 = \overline{w}$ as a new variable in equations (20). The resulting equations will then, in general, contain a term in $\overline{w}$. Assume some non-zero solution $\phi(u, v)$ of the equations in $\overline{w}$ and then carry out the substitution $\overline{w} = \phi(u, v)w$. Since the equations are homogeneous, they will then lack the term in w.

$$
\begin{aligned}
w_{111v} &= aw_{112} + bw_{122} + cw_1 + dw_2 \\
w_{112v} &= a'w_{112} + b'w_{122} + c'w_1 + d'w_2 \\
0 &= w_{11} + a''w_{12} + c''w_1 + d''w_2.
\end{aligned} \tag{21}
$$

We see that if there is to be no second parabolic solution in the system resulting from the integrability conditions, b must equal zero. Assume that $d'' \neq 0$, for if d'' in the third equation of (21) is zero, a second parabolic solution may be obtained directly by differentiating this equation with respect to v and eliminating the Z_{112} and Z_{122} terms by means of the first two

equations of (21). This leaves all first and second derivative terms with no Z_{11} term present so that another differentiation brings in the other parabolic solution.

If now we differentiate the third equation of (21) with respect to u and by means of this original third equation eliminate the Z_2 term from the first two of (21) and from the equation obtained by differentiation we obtain:

$$
\begin{aligned}
w_{112} &= fw_{11} + gw_{12} + hw_1 \\
w_{122} &= f'w_{11} + g'w_{12} + h'w_1 + b'w_{22} \\
w_{111} + k''w_{112} &= f''w_{11} + g''w_{12} + h''w_1.
\end{aligned} \tag{22}
$$

$\overline{w}_1 = \overline{W}$ may now be taken as a new variable in the first and third equation of (22). We also may eliminate the $\overline{W}$ term by again assuming a non-zero solution and making the substitution $\overline{W} = \psi(u, v)W$. This results in two second order equations:

$$
\begin{aligned}
W_{11} &= mW_1 + nW_2 \\
W_{12} &= m'W_1 + n'W_2,
\end{aligned} \tag{23}
$$

which have for the general solution: [6]

$$
\begin{aligned}
W &= [Ag(u, v) + B(v)] \\
W &= [B(v) + g(u, v)B'(v)],
\end{aligned} \tag{24}
$$

where $g(u, v)$ is a known function, $B(v)$ an arbitrary function of v, and A an arbitrary constant.

By multiplying equations (24) by $\psi(u, v)$ and integrating once with respect to u, we have the two possible forms of the solution w of the first and third equations of (22):

$$
\begin{aligned}
w &= g_1(u, v)A + g_2(u, v)B(v) + F(v) \\
w &= g_1(u, v)B(v) + g_2(u, v)B'(v) + F(v),
\end{aligned} \tag{25}
$$

where the g's are again known functions and the $F(v)$ an arbitrary function of v.

[6] L. c., p. 32.

It is now necessary to make equations (25) satisfy the second equation of (22) and from this condition we obtain the restrictions on $F(v)$. b' in this second equation of (22) cannot be equal to zero or a second parabolic would consequently be implied.

Thus, taking the first equation of (25) and, after carrying out the indicated differentiation, we have upon substituting in the second equation of (22):

$$F'' = g_3(u, v)A + g_4(u, v)B(v) + g_5(u, v)B'(v) + g_6(u, v)B''(v).$$

As $F(v)$ and $B(v)$ are functions of v alone and as $B(v)$ is completely arbitrary, this relation has to hold for all values of u. Therefore, place $u = 1$.

$$F'' = g_3(v)A + g_4(v)B(v) + g_5(v)B'(v) + g_6(v)B''(v). \tag{26}$$

Now let us assume that $F(v)$ and $B(v)$ can be expressed as a linear combination of the arbitrary constant, the arbitrary function and its derivatives. That is,

$$\begin{aligned} F(v) &= aA + bL + cL' \\ B(v) &= dL + fL', \end{aligned} \tag{27}$$

where a, b, c, d, and f are undetermined functions of v.

If now the indicated derivatives of (27) are found and substituted in (26) we obtain—by equating coefficients of the arbitrary constant, the arbitrary function and its derivatives—five total differential equations which determine the five functions a, b, c, d, and f. Thus we have shown that $F(v)$ and $B(v)$ can be expressed as a linear combination of A, L and L'.

The process also can be reversed. From the first equation of (27), for a given $F(v)$ and A, we may determine L from the solution of the total differential equation. If, then, also the first term of $B(v)$ be taken as dL, which is permissible, then the next term must be a function times L' unless the function is zero, or if the function is zero there is the next derivative times a new function which constitutes the next term of $B(v)$ unless it too is zero, etc.

In the same way, using the second equation of (25) we are able to express $F(v)$ and $B(v)$ as follows:

$$F(v) = aL + bL' + cL''$$
$$B(v) = dL + eL'.$$

Thus we have as a final result the two equations:

$$w = g_7(u, v)A + g_8(u, v)L(v) + g_9(u, v)L'(v)$$
$$w = g_7(u, v)L(v) + g_8(u, v)L'(v) + g_9(u, v)L''(v).$$

If some of the coefficients of (26) are zero, the reasoning is slightly modified, but when account of the original system is taken, no new forms result except in the case where g_4 and g_5 are zero and $g_8 = 1$, then we may have:

$$F(v) = L + cA$$
$$B(v) = L + D.$$

where D is an arbitrary constant.

Thus we have for the three solutions which satisfy the three equations of (23) as

$$\begin{aligned} w &= g_7(u, v)A + g_8(u, v)D + g_9(u, v)B(v) \\ w &= g_7(u, v)A + g_8(u, v)B(v) + g_9(u, v)B'(v) \\ w &= g_7(u, v)B(v) + g_8(u, v)B'(v) + g_9(u, v)B''(v). \end{aligned} \tag{28}$$

They must also have for the solution $w = 1$, so that g_9 in the first equation, g_8 in the second equation and g_7 in the third must equal one.

If we now multiply the equations (28) by $\phi(u, v)$ and integrate once with respect to u, we have:

$$\begin{aligned} Z &= h_1(u, v)A + h_2(u, v)D + h_3(u, v)B(v) + F(v) \\ Z &= h_1(u, v)A + h_2(u, v)B(v) + h_3(u, v)B'(v) + F(v) \\ Z &= h_1(u, v)B(v) + h_2(u, v)B'(v) + h_3(u, v)B''(v) + F(v), \end{aligned} \tag{29}$$

where the $F(v)$ is an arbitrary function of v, and the h's definite determined functions of both variables.

If now we eliminate the z_2 term from the third equation of (18) by means of either of the first two and then proceed exactly as before to determine the $F(v)$ in (29) by making them satisfy this equation, we have finally:

$$
\begin{aligned}
Z &= g_1(u, v)A + g_2(u, v)D + g_3(u, v)E + g_4(u, v)B(v) \\
Z &= g_1(u, v)A + g_2(u, v)D + g_3(u, v)B(v) + g_4(u, v)B'(v) \\
Z &= g_1(u, v)A + g_2(u, v)B(v) + g_3(u, v)B'(v) \\
&\qquad + g_4(u, v)B''(v) \\
Z &= g_1(u, v)B(v) + g_2(u, v)B'(v) + g_3(u, v)B''(v) \\
&\qquad + g_4(u, v)B'''(v),
\end{aligned} \tag{30}
$$

where E is an arbitrary constant.

Again, these equations must be satisfied by $Z = 1$, so that the coefficients of the $B(v)$ terms are equal to one.

If these solutions are to satisfy (1) they must be multiplied through by the $\phi(u, v)$ which eliminated the Z term from these original equations.

Thus we finally have as the solution of (1) where only one parabolic is implied:

$$
\begin{aligned}
Z &= \phi(u, v)[g_1(u, v)A + g_2(u, v)D + g_3(u, v)E + B(v)] \\
Z &= \phi(u, v)[g_1(u, v)A + g_2(u, v)D + g_3(u, v)B'(v) + B(v)] \\
Z &= \phi(u, v)[g_1(u, v)A + g_2(u, v)B'(v) + g_3(u, v)B''(v) \\
&\qquad + B(v)] \\
Z &= \phi(u, v)[g_1(u, v)B'(v) + g_2(u, v)B''(v) \\
&\qquad + g_3(u, v)B'''(v) + B(v)]
\end{aligned} \tag{31}
$$

on the assumption that c'' and either f or f' in (18) are not zero.

7b. *Special Case*

We shall now consider the cases previously excluded by the assumptions on c'', f, f'.

First, suppose f and f' are both equal to zero in (18), but c'' is not equal to zero. Then $\overline{w} = Z_1$ may be taken as a new variable in the first two equations and the term in $\overline{w}$ removed by assuming a solution. Then these two second order equations

in $\overline{w}$ will have the solution (24). If we then multiply through by the assumed solution and integrate once with respect to u, we have the solutions Z satisfied by the first two equations of (18). If we then follow the general method as before and make the solutions satisfy the third equation with $c'' \neq 0$, we have the final solution in the form of equations (28).

Second, suppose f, f' and c'' all $= 0$, then f'' cannot be equal to zero or a second parabolic would be implied in the system directly. If f'' is not equal to zero, we may again make our solutions satisfy this third equation and by taking our two arbitrary functions as slightly different linear combinations of the third, we again arrive at (28) for the solution. In each of these cases the coefficient of the L term must be one due to the fact that $w = 1$ is a solution.

Third, let us assume that f and f' are not both zero, but $c'' = 0$.

f'' cannot be equal to zero or again the other parabolic would be implied. Thus we may use one of the equations to eliminate the Z_2 term from the remaining two. Then if $Z_1 = w$ is taken as a new variable and the w term removed by introducing a solution, we have two general second order partial differential equations. Since one parabolic solution exists in the system, we may take it as a new variable and thus reduce the general second order equation to (23). This equation when multiplied through by the solution and integrated, will give us finally a set of solutions similar to (28) where the coefficient of the L term is again one.

Thus all the cases previously excluded lead to results which are only special cases of the results in (31).